

Article

# Using Data Mining in Educational Administration - A Case Study on Improving School Attendance

Raymond Moodley <sup>1</sup>, Francisco Chiclana <sup>1,2</sup>, Jenny Carter <sup>3</sup> and Fabio Caraffini <sup>1\*</sup>

<sup>1</sup> Institute of Artificial Intelligence, De Montfort University, Leicester, UK;

raymond.moodley@my365.dmu.ac.uk, chiclana@dmu.ac.uk, fabio.caraffini@dmu.ac.uk

<sup>2</sup> Andalusian Research Institute on Data Science and Computational Intelligence (DaSCI), University of Granada, 18071 Granada, Spain

<sup>3</sup> Department of Computer Science, University of Huddersfield, Queensgate, Huddersfield, UK; j.carter@hud.ac.uk

\* Correspondence: fabio.caraffini@dmu.ac.uk

Version April 27, 2020 submitted to Appl. Sci.

**Abstract:** Pupil absenteeism remains a significant problem for schools across the globe with its negative impacts on overall pupil performance being well-documented. Whilst all schools continue to emphasize good attendance, some schools still find it difficult to reach the required average attendance, which in the UK is 96%. A novel approach is proposed to help schools improve attendance that leverages the market target model, which is built on association rule mining and probability theory, to target sessions that are most impactful to overall poor attendance. Tests conducted at Willen Primary School, in Milton Keynes, UK, show that significant improvements can be made to overall attendance, attendance in the target session, and persistent (chronic) absenteeism, through the use of this approach. The paper concludes by discussing school leadership, research implications, and highlights future work which includes the development of a software program that can be rolled-out to other schools.

**Keywords:** Educational Data Mining; Association Rule Mining; Improving School Attendance; Persistent Absenteeism.

## 1. Introduction

Pupil attendance remains a key focus for schools, local authorities and national governments across the world as a result of its strong, positive correlation with pupil attainment, pupil well-being and improved economic outcomes for pupils later in life [1–3]. In the UK, the Department for Education (DfE) has strict policies on school attendance with legal obligations for both parents, which also includes guardians in this study, and schools [4]. Parents are legally obliged to send their children to school and ensure regular attendance, while schools have a legal duty to take the necessary steps and have policies in place to effectively manage pupil attendance [4]. In this regard, there is a significant requirement from schools to be proactive on attendance management as they must: accurately record attendance, proactively follow-up with parents on all absences, and put initiatives in place to manage and encourage good attendance [4].

As further clarified in section 2.2.1, the underlying reasons as to why pupils are absent from school have been well-studied and generally fall into one, or more, of three categories being 1) unable to attend school due to other obligations (e.g. illness, carer duties, family instability); 2) avoiding school due to fear, embarrassment, boredom (e.g. being bullied) and 3) pupil/family do not place value in schooling and/or have other activities that they would rather do, e.g. taking a vacation, or high levels

31 of illiteracy within the family [1,5]. To this end, strategies for managing absenteeism (predominantly  
32 qualitative) have also been well-studied with models, frameworks and initiatives for improving school  
33 attendance being proposed and evaluated [1–3,5]. The quantitative approaches involving school  
34 attendance has primarily been seen as a task within the Educational Data Mining (EDM) branch of  
35 research, where the key objective is to improve pupil performance through the use of data mining and  
36 artificial intelligence (AI) techniques [6]. Indeed, there have been several models proposed to predict  
37 pupil outcomes, however, attendance has typically been used as an input variable for these models as  
38 opposed to being a key focus area [6–10].

39  
40 The case for increased use of data analytics and AI to improve attendance has been well-made,  
41 however very little use-cases and readily available analytics models exist that can be easily adopted by  
42 school practitioners to improve school attendance [1,2]. Further, most school practitioners are new  
43 to data analytics and have no previous data science background. Despite the current availability of  
44 training and certification courses, it is often challenging for practitioners to develop their models  
45 and algorithms to conduct a deep analysis of data [11]. It is against this backdrop that this study  
46 aims to provide school practitioners with a simple, yet effective model to improve school attendance  
47 by identifying and acting on attendance patterns that are not obvious to extrapolate without data  
48 analytic skills. The proposed model was applied to a live setting using Willen Primary School (WPS), a  
49 local authority-maintained primary school in Milton Keynes, UK, as a case study. In principle, the  
50 approach, findings and recommendations from this study can be leveraged by other schools wanting  
51 to improve pupil attendance. In this regard, intervention programs may have to be adapted to cater  
52 for the school's specific circumstances.

53  
54 The remainder of the paper is divided as follows: An overview of the relevant literature is provided  
55 in Section 2 followed by a detailed description of the problem statement and development of the  
56 underlying analytical model in Section 3. The research methodology is outlined in Section 4, with  
57 a presentation of the results and discussion in Section 5. Finally, conclusions and future work are  
58 detailed in Section 6.

## 59 2. Literature Review

### 60 2.1. Educational Data Mining

61 The definition of EDM in [11] accurately surmises the approach of using what was  
62 once commercial data mining techniques to improving outcomes in education, including  
63 government-sponsored education. EDM, according to [11], “seeks to analyse educational data  
64 repositories to better understand learners and learning and to develop computational approaches that  
65 combine theory and data to transform practice to benefit learners”. Similar definitions for EDM were  
66 provided in [12] and [13] where EDM is defined as a knowledge extraction process where valuable  
67 insights are obtained from data originating from an educational setting. In this regard, EDM may  
68 be compared to commercial techniques like Market Basket Analysis (MBA), which is in essence, a  
69 technique that leverages data analytics on customer transaction data to enhance customer engagement,  
70 and transaction intensity within the retail sector [6,14–16].

71  
72 The popular MBA techniques of Clustering and Association Rule Mining (ARM) have been widely  
73 used in EDM in a variety of contexts. Daniel, in [14], Merceron et al., in [15], and Weng, in [13],  
74 noted that ARM has been very useful in educational applications such as: finding mistakes that are  
75 commonly made together by students, making recommendations to students on e-learning course  
76 choices, and finding associations in behavioural patterns of students. Similarly in [17], ARM was  
77 used to find factors that influenced student performance in courses, with the study concluding that  
78 student performance was directly correlated to attention in class (including attendance), completing

79 assignments and good note-taking.

80  
81 Clustering has also been widely used in education with success. In their review of clustering within  
82 EDM, Dutt et al., in [18], discussed the various educational contexts in which clustering was used  
83 including: using K-Means clustering to improve learning by grouping students with similar learning  
84 styles; and clustering brain scans of students who showed similar responses to learning into groups  
85 and targeting each group differently to improve learning. Similarly, clustering was also used to  
86 understand student behaviour in online learning environments by comparing sequential student data  
87 and leveraging a clustering algorithm to group like-minded students [19]. It should be noted that while  
88 clustering does have its place, it needs to be done carefully within the government schooling sector as  
89 it may be perceived by some parents as unfairly “targeting” groups of pupils, which is generally not  
90 the case [20].

## 91 2.2. School Absence

92 There exists a myriad of terms used to describe school absence which helps focus diagnoses so  
93 that targeted plans could be put in place to address their underlying causes [1,3,4,21]. While some  
94 absences may be seen as acceptable, in the UK, schools have become tough on all absences irrespective  
95 of their reason, as they are equally destructive to learning [1,4]. Authorised absence, defined as an  
96 acceptable absence approved by the school (e.g. illness or bereavement) is typically granted, but  
97 schools have become wary of its abuse, particularly close to ending of term, when parents want to  
98 capitalise on cheaper holidays without incurring fines [4,22,23]. On the other hand, unauthorised  
99 absence (absent without permission) has received widespread condemnation from lawmakers and  
100 education non-profit organisations, with several cases being trailed in court, or parents being fined in  
101 line with the local authority and national government policy [4,22,23].

102  
103 The concepts of school refusal (SR) and truancy form part of unauthorised absence and has been  
104 well-outlined in [3] and [21], with SR defined as non-attendance due to the expectance of strong  
105 negative emotions while at school (e.g. fear as a result of bullying, embarrassment as a result of being  
106 teased or separation anxiety), while truancy is related to anti-schooling sentiments (without parental  
107 consent) including finding school boring or finding activities outside of school more attractive (e.g.  
108 going to the cinema during school time). School withdrawal (e.g. taking time off to go on holiday)  
109 is similar to truancy but with parental consent, and is generally very difficult to address once it  
110 becomes excessive as it usually requires multi-agency involvement that focuses on the family as well  
111 as the pupil [5]. The notion of persistent absence or chronic absence has also been well-studied, with  
112 the definition in the UK being: where a pupil is absent from school for 10% or more, irrespective  
113 of the reason [1,22]. Persistent absenteeism is being well-tracked by schools and local authorities  
114 in the UK, with initiatives and policies put in place to deal with the problem as it arises [4,22].  
115 However, the situation is not the same in some other developed countries, and is often overlooked  
116 and wreaks havoc long before the problem is diagnosed [1]. In the U.S. for example, Balfanz and  
117 Byrnes, in [1], noted that chronic absenteeism is largely unmeasured and hence not noticed. The  
118 authors further point out that only a few states and cities in the U.S. measure chronic absenteeism,  
119 and even when it is measured, the metric of average daily attendance for the entire school “masks  
120 more than it reveals”. Left unchecked, chronic absenteeism eventually leads to a disengagement  
121 with education and results in poor career prospects for the pupil, and most likely a future of poverty [1].

122  
123 Separation anxiety or in its more severe form, Separation Anxiety Disorder (SAD), is a type of school  
124 refusal and has been well-documented in [24]. SAD is common among young children (up to 1 in 20  
125 children suffer from SAD) and is defined as the fear of leaving the safety of parents or caregivers [24].  
126 Children experiencing SAD often present with tantrums, panic attacks or bad behaviour and can have  
127 a significant negative impact on the child’s academic, social and physiological development [24,25].

128 Indeed, separation anxiety is most common after children have spent long spells with their parents or  
129 caregivers and is common after weekends or holidays, and may also present every morning in some  
130 children after they have spent the previous afternoon and night with parents or caregivers [24,25].

### 131 2.2.1. Why are pupils absent?

132 There is broad consensus by researchers as to why pupils do not attend school, and the underlying  
133 causes for absence fall into three categories [1,5] which are indeed very large by themselves:

- 134 1. unable to attend school due to other obligations;
- 135 2. avoiding school (school refusal);
- 136 3. pupil/ family do not place value in schooling (and/ or have other activities that they rather do).

137 This notion of not placing value in schooling has been further separated into *truancy*, i.e. pupils  
138 staying away from school without parental knowledge, and school withdrawal, also known as parental  
139 condoned absence, i.e. parents condone the absence as it proves beneficial to them or the family at  
140 large [21]. Given the vast array of underlying causes, researchers have tended to become more specific  
141 in examining the problem of absence. In [3] the focus was on school refusal and truancy with peer  
142 relationships and classroom management by teachers as underlying causes. In this regard, Havik  
143 et al. [3] found that both good peer relationships and effective classroom management had strong  
144 positive correlations with good attendance. Similarly, tackling truancy and parental beliefs (as part of  
145 school withdrawal) were the key focus areas of in [2] and [5] respectively, with both studies showing  
146 that there is a strong positive correlation with good attendance and effective, regular communication  
147 between school and home.

### 148 2.2.2. Impacts of absence

149 Balfanz and Byrnes, in [1], were firm in their conclusions that “missing school matters”, noting  
150 that in the US, missing school impacted academic achievement irrespective of age and that those that  
151 were from low-income backgrounds were more impacted by absence as they were less likely to have  
152 provisions at home to make up for the lost time. In the UK, similar sentiments were echoed in [4] and  
153 [22] with respect to absence, including more long term impacts on the pupil, such as social anxiety  
154 and lack of self-confidence, both of which are known pre-cursors to interrupted employment and  
155 consequently lower economic attainment in adulthood [1,21,25]. Whilst these are all significant impacts  
156 in their own right, the key impact of absence, which was noted across several studies including in  
157 [1,2,4,22] and [5], was the long term disengagement with education which not only impacted the pupil  
158 in adulthood but also created the foundation for a vicious cycle when these pupils become parents and  
159 project their negative attitudes towards education onto their children.

### 160 2.2.3. Improving attendance

161 The conceptual framework proposed in [26] for designing interventions to improve attendance  
162 is both relevant and very useful. The proposed three-tier framework targeted all pupils along the  
163 absenteeism spectrum with tier 1 strategies focussed on pupils with emerging attendance problems,  
164 whilst tier 2 focussed on pupils that are at risk of being persistently absent, and tier 3 on those that are  
165 already persistently absent. The overall approach of this framework emphasizes early identification  
166 and treatment, rather than a sole focus on those that are already persistently absent.

167  
168 This approach is well-recognised and several studies have operationalised this framework in varying  
169 depths [1,2,4,22]. In [4] and [22], which are relevant to the UK context, guidelines suggest that all  
170 absenteeism should be tackled with context-specific approaches that include using data analytics,  
171 working with parents, using incentives, and enforcing fines. Similarly, the Early Truancy Prevention  
172 Program (ETPP) introduced in [2] proposed a five-step approach, all of which required the teacher to  
173 be proactive, and work actively with parents to drive-up attendance. Pilot tests using the ETPP did

174 show a significant improvement in attendance [2], however, most initiatives were time-intensive and  
 175 required teachers and school administrators to spend a large amount of time working with parents on  
 176 an ongoing basis. This is not practical in the UK, because teachers are already stretched, and school  
 177 budgets are being squeezed [27]. Efforts to improve attendance in [1] were underpinned by offering  
 178 both short-term and long term rewards through local and national/ state campaigns. At a local level,  
 179 schools offered rewards for pupils who attended regularly that were more meaningful to pupils and  
 180 included fun activities like dance and diplomas for completing short courses. While at a national level,  
 181 school attendance was stressed by senior political figures and “success mentors” who were largely  
 182 celebrities that attributed their success to regular school attendance [2].

### 183 3. Problem Statement and Analytical Model

#### 184 3.1. Problem Statement

185 It is well-documented that providing pupils with the right incentives to attend school results in  
 186 improved attendance, and consequently improved pupil attainment and progress [1,4]. Given this, the  
 187 problem being addressed by this study may be stated as follows: Let  $S$  be a school with all its pupils,  
 188  $U$ . Let the school week,  $J$ , be divided into  $m$  distinct sessions,  $J_i$ , such that  $J_i \in J = \{J_1, J_2, \dots, J_m\}$ .  
 189 Further, let  $T$  be a database in  $S$ , that contains the attendance records of all pupils across all sessions for  
 190 a period,  $W$ . Hence, there may exist a database,  $T_t$ , where  $T_t \subseteq T$ , that contains the attendance records  
 191 of pupils  $U_t$ , where  $U_t \subseteq U$ , who have below the required attendance in at least one school session  
 192 and/or the overall average attendance, but where attendance in all other sessions are above or equal  
 193 to the requirement. In the UK, the required attendance target is 96% [4]. Given that the leadership  
 194 and staff of the school  $S$  are intent on maximising pupil attendance (with the focus on driving up the  
 195 overall average pupil attendance through incentives and interventions) while minimising effort and  
 196 associated costs (largely incentives and staff costs), it becomes necessary to optimise the targeting of  $J_i$ .  
 197 Thus, this study aims to provide a framework, and useful tool for schools, based on ARM and Frequent  
 198 Itemset Mining (FIM), for targeting the right school session(s) with incentives and interventions that  
 199 maximises the impact on improved overall school attendance.

#### 200 3.2. Analytical Model

201 We commence by noting the definitions of the well-known ARM concepts of support, confidence,  
 202 minimum support, minimum confidence, and the Apriori principle first introduced in [28], and as  
 203 detailed in [29] and [30].

- The support of an item  $A$ , in a transaction database  $T$ , is given by:

$$\text{supp}(A) = P(A) = \frac{\text{number of transactions in } T \text{ that contain } A}{\text{number of transactions in } T}$$

- The probability of the presence of item  $A$  leading to the presence of item  $C$  (commonly referred to as confidence) is given by:

$$\text{conf}(A \rightarrow C) = \frac{\text{number of transactions in } T \text{ that contain both } A \text{ and } C}{\text{number of transactions in } T \text{ that contain } A}$$

204 When  $\text{supp}(A)$  exceeds some user-defined value for support (commonly referred to as minimum  
 205 support or  $\text{minsup}$ ) we note that  $A$  is considered to be frequent. Similarly when  $\text{conf}(A \rightarrow C)$  exceeds  
 206 some user-defined value for confidence (commonly referred to as minimum confidence or “ $\text{minconf}$ ”)  
 207 we note that  $A$  and  $C$  are considered to be associated. Note that FIM is defined as the process of  
 208 finding all itemsets that exceed  $\text{minsup}$  in a given database [13,16,29].

210 The Apriori principle, first detailed in [28], and more recently in [16], states that for a given set  
 211 of transactions,  $\text{supp}(A) \geq \text{supp}(A, C)$ . This is consistent with probability theory where  $P(A) \geq$   
 212  $P(A \cap C)$ , as well as in practical terms, e.g. where the number of transactions that contain pupils who  
 213 are absent on Monday AM is always greater than or equal to the number of transactions that contain  
 214 absences on both Monday AM and PM.

### 215 3.2.1. Identifying the Best Sessions to Target with Attendance Improving Initiatives

216 Pupils that have above or equal to the required attendance in every session are generally  
 217 considered to have very good attendance, and in essence help the school boost its overall average  
 218 attendance. Let  $T_p$  be a database containing the attendance records of all pupils that are persistently  
 219 absent, hence  $T_p \subseteq T_t$ . Persistent absenteeism in the UK is defined as having an overall average  
 220 attendance of less than 90% [22]. Given that schools take severe action once attendance drops below  
 221 85%, including removing a pupil from the school roll,  $T_t$  thus represents a significant portion of  $T$   
 222 for a school that has overall below-the-required-average attendance [4]. Hence, improving pupil  
 223 attendance in  $T_t$  will enhance overall attendance, and as most schools have limited resources, the  
 224 question of which  $J_i$  in  $T_t$  should be targeted often arises. Intuitively, the best session to target should  
 225 be that session which has both the highest absence and the highest association with poor overall  
 226 average attendance,  $O$ . This scenario may be represented in terms of ARM as targeting the session  
 227 where  $\text{supp}(J_i)$  and  $\text{conf}(J_i \rightarrow O)$  is the largest. However, we also note that scenarios do exist where  
 228  $\text{supp}(J_c) > \text{supp}(J_k)$  but  $\text{conf}(J_c \rightarrow O) < \text{conf}(J_k \rightarrow O)$ . In these cases, the choice between  $J_k$  and  $J_c$   
 229 is not obvious.

230  
 231 This choice-making problem is not unique to school attendance and often arises in several other sectors  
 232 including in retail, medicine, and security [16]. We note that a similar problem involving the selection  
 233 of the best item to target for grocery retail promotions has recently been addressed in [16], and thus the  
 234 methods employed in that study could be applied here. To facilitate easy processing,  $T_t$  is converted  
 235 into a database with binary attributes, with sessions and/or the overall average attendance being  
 236 assigned a “1” when attendance drops below the required levels. Clearly,  $T_t$  may now be considered to  
 237 be an absenteeism database.

### 238 3.2.2. Applying the Market Target (mt model) on school attendance data

239 The mt model proposed in [16] was shown to be effective in making choices between items in the  
 240 form  $(A \rightarrow C)$  and  $(B \rightarrow D)$ . Indeed, the problem laid out in Section 3.2.1 is of the form  $(A \rightarrow C)$  and  
 241  $(B \rightarrow C)$ , and may be considered a subset of the more generalised choice making problem that the mt  
 242 model addresses.

243  
 244 Let  $P(J_i)$  be the support of session  $J_i$  in database  $T_t$ , and  $P(J_i, O)$  be the support of session  $J_i$  and  $O$   
 245 co-occurring in database  $T_t$ . In practical terms,  $P(J_i, O)$  may be viewed as the number of children,  
 246 or instances, that have both below the required attendance for  $J_i$  and  $O$  in the database  $T_t$ . Thus by  
 247 definition,  $\text{conf}(J_i \rightarrow O) = P(J_i, O) / P(J_i)$ . As was the case in grocery retail, detailed in [16], there  
 248 are two intuitive schools of thought on solving this problem to reduce attendance. One may suggest  
 249 targeting the session,  $J_i$ , that has the highest  $\text{conf}(J_i \rightarrow O)$ , as a reduction in every absenteeism in  $J_i$   
 250 will most likely lead to a reduction in  $(J_i, O)$ . However, if  $P(J_i, O)$  is low, then  $(J_i, O)$  may be considered  
 251 to be rare, and solving this scenario may not have the desired overall impact on  $O$ . Rare rules, as  
 252 defined in [13], are rules that are highly associated but occur less frequently in a dataset, i.e. they  
 253 have lower support. Conversely, targeting a high  $P(J_i, O)$  may seem attractive, but if  $\text{conf}(J_i \rightarrow O)$   
 254 is low, then lowering  $P(J_i)$ , through some initiatives, may not have the required impact on  $P(J_i, O)$ ,  
 255 and consequently  $P(O)$ . Thus, it is evident that a model that takes into consideration the concepts of  
 256 support and confidence is required to find the optimum solution. In this regard, the mt model, detailed  
 257 in [16], is a model that addresses this exact challenge.

### 258 3.2.3. Adapting the mt model for school attendance

259 The mt model, adapted for school attendance, is developed below, and in essence evaluates  
 260 options, for example: option  $(J_i, O)$  and  $(J_k, O)$ , based on both the support and confidence of that  
 261 option in the database. It is clear that  $P(J_i) \geq P(J_i, O)$  for all  $i \in m$ , hence the underlying principle of  
 262 the mt model is that it evaluates the “effort” required to make  $P(J_i, O) = \text{minsup}$ , which is considered  
 263 to be the “desired” state of  $P(J_i, O)$ . Note that in this instance, the “desired” state is equivalent to the  
 264 maximum session and overall absenteeism. Also note that minsup is user-defined, and is governed by  
 265 the Apriori principle, i.e.  $P(J_i, O) \leq \text{minsup} \leq P(J_i)$  for all  $i \in m$ . The number of absences required for  
 266 a  $P(J_i, O)$  combination to reach the “desired” state is given by Equation (1), where  $|T_t|$  is the number  
 267 of transactions in database  $T_t$ .

$$\text{Number of absences required for “desired” state} = (\text{minsup} - P(J_i, O)) \cdot |T_t| \quad (1)$$

268 Given that not all children absent in  $J_i$  will also be absent in  $O$ , the “market target” referred to [16], or  
 269 in this instance, the pupil target, may thus be defined as the number of required absences in  $J_i$  such  
 270 that the number of absences required for the “desired” state of  $(J_i, O)$  in  $T_t$  to be reached. This is stated  
 271 mathematically in Equation (2), where  $P(J_i, O)/P(J_i) = \text{conf}(J_i \rightarrow O)$ .

$$\text{pupil target} \cdot \frac{P(J_i, O)}{P(J_i)} = \text{Number of absences required for “desired” state} \quad (2)$$

272 Thus the mt equation, given in Equation (3) is obtained by combining Equations (1) and (2), and  
 273 dividing both sides the minimum support, i.e. the physical number of absences equivalent to minsup  
 274 in database  $T_t$ . Note that mt is a normalised parameter, and is given by pupil target / minimum  
 275 support.

$$\text{mt} = \frac{P(J_i)}{P(J_i, O)} - \frac{P(J_i)}{\text{minsup}} \quad (3)$$

276 From Equation (3), it is evident that options that have the lowest mt value require the lowest “effort” to  
 277 reach the “desired” state, and are thus considered the best choices for a given minsup. From a practical  
 278 perspective, this implies that the school targets the school session that has the greatest propensity to  
 279 lead to overall below-average school attendance. There are also practical constraints of managing a  
 280 school that must be considered. In this regard, initiatives must target all, or the majority of school  
 281 pupils to ensure fairness, and given that most initiatives are largely fixed costs (e.g. the effort in  
 282 planning activities is similar whether the audience is 100 or 250), it makes sense to target the session  
 283 which impacts overall absenteeism the most [20]. The most impactful session is the session which has  
 284 the lowest mt value as it requires the least “effort” to reach the “desired” state.

### 285 3.2.4. Algorithm for identifying target sessions using the mt model

286 Applying the mt model to identify the best sessions to target is relatively straightforward. The  
 287 mt value is computed for each  $(J_i \rightarrow O)$  combination and the one with the lowest mt value is the best  
 288 session to target. The steps of the proposed algorithm are detailed in Algorithm (1).

## 289 4. Experiments

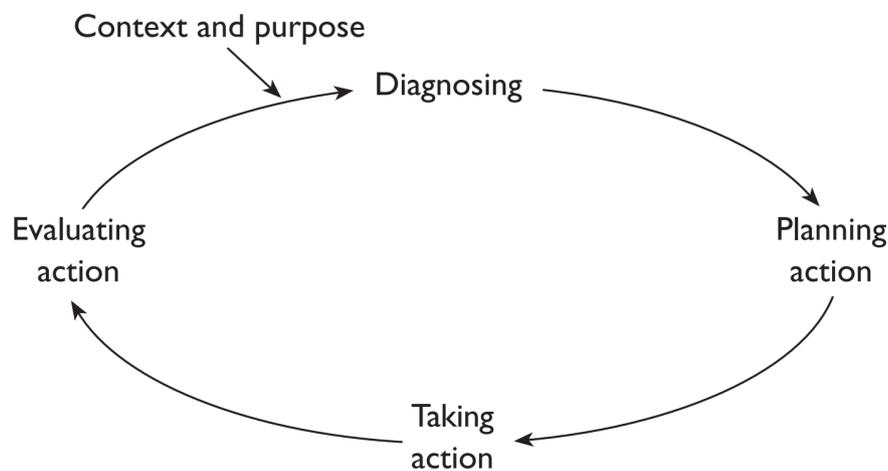
### 290 4.1. Experimental Process

291 Experiments were conducted based on the well-known Action Research process as detailed in  
 292 [31] and outlined in Figure 1. As per [31], the process begins by defining the context and purpose by  
 293 asking the question: why is this project required or desirable? However, it is the diagnosing phase that  
 294 usually proves to be the most challenging as it involves identifying the possible issues or the most  
 295 impactful issue, which is sometimes not obvious [32]. Consequently, data analytics is often leveraged

**Algorithm 1:** Identifying target sessions using the mt model

- 1 Create the dataset,  $T_t$ , from  $T$  that contains the attendance records of all pupils that are on the roll for the entire period, and where their attendance has been below the required level in at least one session
- 2 Using a ARM/FIM algorithm (e.g. Apriori or ECLAT) with a low support and confidence, find  $\text{supp}(J_i)$  and  $\text{conf}(J_i \rightarrow O)$  for all sessions
- 3 Calculate the mt value for each  $(J_i, O)$  combination using an appropriate value for minsup
- 4 Order sessions based on mt values, with the session that has the lowest mt value being the best session to target

296 to simplify this task through the use of models and algorithms to process data into information [32]. In  
 297 this regard, the mt model forms part of the diagnosing phase. Note that the Action Research process is  
 298 cyclical and actions taken have to be regularly evaluated against the context and purpose, which could  
 299 also change over time [31].



**Figure 1.** Action Research Process as outlined in [31]

300 Research was conducted at Willen Primary School with the context and purpose of improving overall  
 301 school attendance to be above or in line with the national requirement, which is currently set at 96%  
 302 in the UK [4,22]. The mt model (part of the diagnosing phase) was then used to identify the session  
 303 which was most impactful to overall school absence. Options for possible action were brainstormed  
 304 with school leadership and evaluated in the planning action stage. Following this, selected actions  
 305 were carried out at the school over several months (taking action stage) with the impact on overall  
 306 school attendance then assessed in the evaluating action stage.

#### 307 4.2. Experimental conditions

##### 308 4.2.1. Willen Primary School

309 WPS is a mixed, 2-form entry primary school on the north-eastern side of Milton Keynes, catering  
 310 for 4 to 11 years old children. The school has a capacity of 420 pupils and had 366 pupils on its roll at  
 311 the end of July 2019, with approximately 35% of its pupils coming from outside the school's catchment  
 312 area [20,33]. The school was rated "Good" by the UK's Office for Standards in Education, Children's  
 313 Services and Skill (Ofsted) in its last inspection, which was conducted in November 2017 [34]. Whilst  
 314 the inspector cited very good attendance management practices by the school leadership, he did note  
 315 that further improvement should be made [34]. Given this, the school has continued to fervently

316 promote the importance of good attendance and explored the use of novel approaches to address the  
317 issue of absenteeism giving rise to this study [20].

#### 318 4.2.2. Diagnosing

319 School attendance data for the previous three academic years, i.e. 2015/16, 2016/17 and 2017/18,  
320 were used as the basis for improving school attendance in 2018/2019. For the sake of completeness,  
321 detailed attendance and school roll data are provided in Appendix A. These data were first scrubbed,  
322 to remove pupils that either joined the school after the start of the academic year or left the school  
323 during the academic year, thus producing a dataset  $T$  for each academic year  $W$ , as further discussed  
324 in Sections 3.2.1 and 3.2.4. Subsequently, data were further filtered to produce  $T_i$  by selecting those  
325 pupils  $U_i$  who either had an overall attendance of less than 96% (the required national average) and/or  
326 who were absent at least three times per session during the academic year. Given that the cardinality  
327 of sessions was generally between 34 and 39 per year, it was not practical to filter these sessions at the  
328 96% level as it was too restrictive (equivalent to two absences per session). It is not uncommon for  
329 some children to have up to two absences for some sessions and still have an overall attendance of  
330 at least 96% [20]. The restriction on analysing pupils that were present for the entire academic year  
331 was placed to ensure that the data analysis process was not unfairly skewed. For example, consider a  
332 scenario that occurs fairly regularly: some pupils may enrol at WPS after the start of the academic year,  
333 have 100% attendance for two weeks and then transfer to another school (possibly one that is closer  
334 to their home) [20]. In this case, these pupils will have 100% attendance and be treated analytically  
335 as the same as pupils who had 100% attendance for the entire academic year. Consequently, these  
336 pupils were excluded from the analysis. The size of  $T_i$  for each academic year is detailed in Appendix A.

337  
338  $T_i$  for each academic year was then analysed using an FIM algorithm in R, with a minsup of 0.3 and a  
339 minconf of 0.3 to prune rare rules, similar to the process outlined in [16]. The value of minsup and  
340 minconf was chosen to be low enough to capture all essential rules but high enough to eliminate  
341 superfluous rules. Given that minsup and minconf are user-defined parameters, the choice of an  
342 appropriate value is typically based on the context. Whilst the values of minsup, and minconf have  
343 practical significance in some sectors and settings, e.g. in grocery retail where it is used to identify  
344 popular products [16], in this context, it is used to simplify the data processing by reducing the number  
345 of rules produced. The choice of 0.3 was based on trial and error, which is typically the case in data  
346 processing applications. An initial test pass on the dataset for the 2017/18 academic year using minsup  
347 = 0.4 resulted in some essential rules being pruned, e.g. Wed-PM and Thur-PM, hence minsup was  
348 adjusted to be lower than 0.4. It should be noted that choosing a value inferior to 0.3 will still achieve  
349 the objective, but will increase processing effort. For completeness, the number of rules extracted per  
350 academic year is detailed in Appendix A. Following this step, the output from the FIM stage was  
351 further analysed, using Microsoft Excel, to compute the  $mt$  value for each frequent itemset from which  
352 the best target session was identified.

#### 353 4.2.3. Planning Action

354 Given that the school has strict obligations, guidelines and its strategic agenda that it must  
355 adhere to, it was realized that a multi-prong approach had to be undertaken with regards to planned  
356 actions that would improve attendance and validate the targeting approach proposed in this study.  
357 These planned actions were over and above what the school was currently doing to monitor and  
358 promote attendance. Hence a two-pronged approach was adopted with 1) session-targeting focused on  
359 demonstrating that session (and overall) attendance can be improved by targeting identified session(s);  
360 2) overall attendance improvement initiatives focused on improving attendance in line with the  
361 strategic and statutory obligations of the school.

362 Several alternatives were considered by school leadership and based on their experience, the best  
363 two selected were: 1) focus on shorter periods with prize-based rewards for full attendance; 2) create

364 more exciting initiatives for targeted sessions. The selected initiatives were consistent with the tiered  
365 approach described in [26].

#### 366 4.2.4. Taking Action

367 Apart from continuing to fulfil its statutory and strategic objectives with regards to attendance  
368 (including dealing with persistent absenteeism, promoting and fostering a good environment for  
369 improved attendance, and dealing with truancy) the school implemented the two initiatives outlined  
370 in Section 4.2.3.

371 Initiative One (I1) focused on increasing the frequency and perceived meaningfulness of the rewards  
372 for full attendance so that pupils could both feel tangibly rewarded for full attendance and know that  
373 they can always be eligible for rewards in the next reward period should they not win in the current  
374 or previous period. I1 commenced at the start of the Spring term in January 2019, with all pupils  
375 that had full attendance for the month placed in a draw to win one of eight tickets to a popular, local  
376 trampoline park. The reward was meaningful to the pupils as it was something that they enjoyed  
377 and it was something that was not always available to them due to cost constraints [20]. Given this,  
378 there was considerable excitement from pupils when the initiative was introduced. Initiative Two (I2)  
379 was geared towards targeting the sessions that had the largest impact on poor attendance. Exciting  
380 activities were conducted during the most impactful session throughout the Summer term starting at  
381 the end of April 2019. These activities, which were centred on a common theme and designed to be  
382 in line with the learning objectives, involved the entire school and included elements that the pupils  
383 would consider exciting [20]. Further details on I2 are provided in Section 5.

#### 384 4.2.5. Evaluating Action

385 Following the implementation of the initiatives, the pupil attendance records for the 2018/19  
386 academic year were analysed using Microsoft Excel and compared with previous years to quantify the  
387 impact of I1 and I2. This then fed into school planning operations for the 2019/20 academic year.  
388 We adopt a simple, inference-based approach by establishing two null hypotheses, and by inference  
389 draw conclusions on the 2018/19 year. The null hypotheses are stated as follows:

- 390 •  $H_0^{(1)}$ : There is no statistically significant change in the attendance data across the three academic  
391 years (2015/16, 2016/17 and 2017/18).
- 392 •  $H_0^{(2)}$ : There is no statistically significant change in the attendance data across all four academic  
393 years.

394 From the above, if  $H_0^{(1)}$  is accepted and  $H_0^{(2)}$  is rejected then we can conclude that the 2018/19  
395 is statistically different from the previous years, and hence the initiatives have made an impact.  
396 Conversely, if both  $H_0^{(1)}$  and  $H_0^{(2)}$  are accepted, then we can conclude that the initiatives have made no  
397 impact on attendance.

## 398 5. Results and Discussion

### 399 5.1. Identifying Target Sessions

400 The average attendance for all pupils who were on the school roll for the entire academic year  
401 was calculated using Microsoft Excel, with the results presented in Table 1.

**Table 1.** Average Attendance for 2015/16, 2016/17 and 2017/18

Session ( $J_i$ )	2015/16	2016/17	2017/18
Mon-AM	93.7%	93.7%	93.8%
Tues-AM	94.1%	94.9%	94.6%
Wed-AM	95.1%	95.0%	95.3%
Thur-AM	94.9%	95.3%	94.8%
Fri-AM	94.6%	94.7%	94.1%
Mon-PM	94.1%	94.5%	94.1%
Tues-PM	94.7%	95.6%	95.1%
Wed-PM	95.5%	95.6%	95.8%
Thur-PM	95.5%	95.7%	95.4%
Fri-PM	94.9%	94.9%	94.5%
Average AM	94.5%	94.8%	94.5%
Average PM	95.0%	95.3%	95.1%
Overall ( $O$ )	94.7%	95.0%	94.8%

402 From Table 1 it can be seen that the school generally did not achieve the required overall average  
 403 attendance of 96% in any of the previous three academic years. Further, attendance in the morning  
 404 (AM) sessions were lower than the afternoon (PM) sessions, with Monday AM being consistently the  
 405 most poorly attended session across the years. This is consistent with theories on separation anxiety  
 406 where young children often dislike going back to school after spending long periods away from school  
 407 with their parents and family, and school withdrawal [22,25]. Separation anxiety may be exacerbated  
 408 when parental collision occurs (school withdrawal) and parents keep pupils at home for fear that they  
 409 may become distressed further [1,21,22].

410

411 Whilst Monday AM is the most absent session based on average percentages, as shown in Table 1,  
 412 the basis of attendance management is not only about increasing the overall average attendance, but  
 413 centred on addressing the most impactful session to overall attendance, which in turn impacts pupil  
 414 performance [1]. It is possible that the most frequently absent session is not the most impactful, as  
 415 children absent in this session could return to school in the next session and have perfect attendance for  
 416 the rest of the week, and generally have good academic performance as well. Thus, any interventions  
 417 aimed at improving attendance in these sessions may likely be less effective as it will be targeting  
 418 children that already have good overall attendance. This may take focus (and valuable resources) away  
 419 from other sessions that may have marginally better attendance, but fraught with problem absenteeism  
 420 that is impacting pupil performance and overall school morale. Indeed, executing “misguided”  
 421 intervention programs can also have detrimental impacts on staff and parents. Staff may lose faith  
 422 in their ability to improve school attendance and performance, and loss morale as their hard work  
 423 may go unrewarded. At the same time parents, whose children are generally good attendees, may  
 424 feel unduly victimised for occasional absences, particularly where such absences are obligatory e.g.  
 425 medical appointments or bereavement [20]. Further, given that these interventions are focused on a  
 426 session that generally comprises of occasional absenteeism, it is unlikely to make a significant impact  
 427 on the children that are chronically absent. Tracking and improving absenteeism, especially chronic  
 428 absenteeism, is a key performance metric of a school’s performance management framework within  
 429 the UK [20,22,34]. As a result, it thus becomes important to target the most impactful session to overall  
 430 attendance, and the use of the mt model, as detailed in Section 3, is one effective way of achieving this  
 431 objective.

### 432 5.1.1. Targeting the Most Impactful Session

433 The mt value for each session was calculated on each  $T_t$  for the previous three academic years,  
 434 as per the process outlined in Section 3, with the results detailed in Tables 2, 3 and 4. Some sessions  
 435 were automatically eliminated, consistent with Lemma 1 in [16], as both their corresponding support

436 and confidence were less than other sessions in the same year. As noted in Lemma 1 in [16] and  
 437 adapted for this study, if both  $\text{supp}(J_i, O)$  and  $\text{conf}(J_i \rightarrow O)$  is less than  $\text{supp}(J_k, O)$  and  $\text{conf}(J_k \rightarrow O)$   
 438 respectively, then  $(J_k, O)$  is the better choice, and  $(J_i, O)$  can thus be eliminated. In Table 2 for the  
 439 2017/18 academic year, Fri-AM had both higher support and confidence than every other session  
 440 except Mon-AM, hence there was no need to compute the  $mt$  value for all other sessions except Fri-AM  
 441 and Mon-AM. The  $mt$  model in Equation (3) was used to decide the better target session between  
 442 Fri-AM and Mon-AM, with a minsup value of 0.550 (the lower support between Fri-AM and Mon-AM)  
 443 being used. Mon-AM had the lower  $mt$  value and hence was selected to be the best session to target.

**Table 2.** Identifying Target Sessions - 2017/2018,  $T_t = 179$

Session ( $J_i$ )	$P(J_i)$	$P(J_i, O)$	$\text{conf}(J_i \rightarrow O)$	$mt$
Mon-AM	0.594	0.561	0.944	-0.021
Mon-PM	0.539	0.517	0.959	-
Tues-AM	0.494	0.472	0.955	-
Tues-PM	0.483	0.472	0.977	-
Wed-AM	0.456	0.433	0.951	-
Wed-PM	0.378	0.356	0.941	-
Thur-AM	0.494	0.467	0.944	-
Thur-PM	0.394	0.383	0.972	-
Fri-AM	0.550	0.539	0.980	0.020
Fri-PM	0.506	0.472	0.934	-
Overall ( $O$ )	0.856	-	-	-

444 The negative value for  $mt$  was also interesting to note. In practical terms, it implied that there were more  
 445 records in  $T_t$  that contained both Mon-AM and  $O$  that were below the required levels than records that  
 446 contained Fri-AM being below the required level. Thus any initiative to resolve absenteeism on Fri-AM  
 447 will always be less impactful than absenteeism on Mon-AM. Hence all other sessions except Mon-AM  
 448 were considered to be rare rules as there exists a  $(J_i, O)$  combination that is under consideration with  
 449  $P(J_i, O) > \text{minsup}$ . This is not always the case and the scenarios were quite different for the 2015/16  
 450 and 2016/17 academic years.

**Table 3.** Identifying Target Sessions - 2016/2017.  $T_t = 180$

Session ( $J_i$ )	$P(J_i)$	$P(J_i, O)$	$\text{conf}(J_i \rightarrow O)$	$mt$
Mon-AM	0.609	0.549	0.903	0
Mon-PM	0.559	0.505	0.904	0.089
Tues-AM	0.505	0.480	0.951	-
Tues-PM	0.436	0.417	0.955	-
Wed-AM	0.490	0.480	0.980	-
Wed-PM	0.456	0.446	0.978	-
Thur-AM	0.480	0.480	1	-
Thur-PM	0.436	0.431	0.989	-
Fri-AM	0.539	0.510	0.945	-
Fri-PM	0.549	0.515	0.938	0.066
Overall ( $O$ )	0.848	-	-	-

**Table 4.** Identifying Target Sessions - 2015/2016,  $T_t = 170$ 

Session ( $J_i$ )	$P(J_i)$	$P(J_i, O)$	$\text{conf}(J_i \rightarrow O)$	mt
Mon-AM	0.624	0.584	0.935	0.009
Mon-PM	0.589	0.558	0.948	0.056
Tues-AM	0.609	0.563	0.925	0.048
Tues-PM	0.569	0.548	0.964	-
Wed-AM	0.492	0.482	0.979	-
Wed-PM	0.467	0.457	0.978	-
Thur-AM	0.558	0.543	0.973	-
Thur-PM	0.487	0.487	1	-
Fri-AM	0.508	0.487	0.960	-
Fri-PM	0.503	0.472	0.939	-
Overall ( $O$ )	0.857	-	-	-

451 From Table 3 for the 2016/17 academic year, all rules except Mon-AM, Mon-PM and Fri-PM were  
 452 shortlisted as the others were determined to be rare. The mt values were computed for each of the  
 453 shortlisted sessions, with minsup set at 0.549 (the lowest support between Fri-PM, Mon-AM and  
 454 Mon-PM). Mon-AM was found to be the most impactful session to overall below-average attendance.  
 455 Similarly from Table 4 for the 2015/16 academic year, Mon-AM was found to be the most impactful  
 456 session with minsup set at 0.589. Given that Monday AM was found to be the most frequent and  
 457 the most impactful session in the three academic years analysed, it can be concluded that the poor  
 458 attendance on Monday AM may be attributed to a combination of school refusal (e.g. due to separation  
 459 anxiety), and school withdrawal/truancy where the return to school may not be seen as being as  
 460 exciting as the weekend that just passed [1]. Therefore, an easier, more exciting start to the school week  
 461 (initiated by the school) may prove successful in addressing this issue.

#### 462 5.1.2. Early Warning System

463 The very high confidence values ( $>0.9$  and in some cases = 1) was also of significant note as it  
 464 suggested that any pupil that was absent for at least three times in any one session was very likely to  
 465 have below overall required attendance. This could be a good tool for the school to use in tackling  
 466 absenteeism as it may be used to identify pupils that are at risk of falling below the requirement,  
 467 consistent with the recommendation in [26]. Further, it could be used as part of conversations with  
 468 parents and pupils in addressing their beliefs and misconceptions about attendance which is consistent  
 469 with the recommendations in [1], [2], and [4] for improving attendance through leveraging analytics.  
 470 This fact-based approach is more likely to resonate well with parents and may negate any possible  
 471 insinuations by parents that their families are being victimised or treated unfairly by teachers and  
 472 school leadership [1,2].

#### 473 5.2. Evaluating the Impacts of Initiatives I1 and I2

474 I1 and I2 were conducted as detailed in Section 4.2.3. Following the results of the analysis  
 475 conducted as part of Section 5.1, the school decided to target Mondays with the emphasis on the  
 476 Monday AM session as part of I2. The Monday Matters initiative was launched in the Summer term  
 477 of 2019 and consisted of a “m-themed” program for five of the ten Mondays during the term. The  
 478 initiatives were selected by the school staff as it represented themes that would resonate well with  
 479 the pupils. The five themed Mondays were: Move-It Monday, Muffin Monday, Mindfulness Monday,  
 480 Mask Monday and Movie Monday. For each themed Monday, pupils were allowed to come to school  
 481 appropriately dressed, e.g. example sports kits on Move-It Monday, and participate in a range of  
 482 planned activities related to that theme which were also linked to the work that was being done in the  
 483 classroom.

## 484 5.2.1. I1: Frequent Rewards for Full Attendance

485 Draws were held every month during the Spring and Summer terms of 2018/19, except for April,  
 486 for all pupils that had full attendance during the month. The April draw was omitted given that April  
 487 had fewer than 10 school days in that month.

**Table 5.** Average Attendance for Spring and Summer Terms: 2015/16, 2016/17, 2017/18 and 2018/19

Session (J <sub>i</sub> )	2015/16	2016/17	2017/18	2018/19
Mon-AM	94.1%	93.9%	94.3%	95.8%
Tues-AM	94.2%	95.2%	95.0%	96.2%
Wed-AM	95.1%	95.5%	95.4%	96.2%
Thur-AM	94.9%	95.5%	95.1%	96.3%
Fri-AM	94.4%	95.0%	94.2%	95.6%
Mon-PM	94.7%	94.6%	94.8%	96.5%
Tues-PM	94.9%	95.8%	95.5%	96.8%
Wed-PM	95.6%	95.8%	95.8%	96.6%
Thur-PM	95.6%	95.8%	95.6%	96.8%
Fri-PM	94.9%	95.1%	94.1%	96.0%
Average AM	94.6%	95.1%	94.8%	96.0%
Average PM	95.2%	95.4%	95.2%	96.5%
Overall (O)	94.9%	95.2%	95.0%	96.2%

488 From Table 5, it is evident that the shorter, more meaningful rewards for full attendance have  
 489 contributed to a significant improvement in overall attendance for the Spring and Summer terms in  
 490 2018/19 with the attendance for every session being considerably higher than the attendance in the  
 491 previous three years. This result was consistent with the findings in [1]

## 492 5.2.2. I2: Monday Matters Initiative

493 Table 6 presents attendance data for Summer term Monday attendance for the 2015/16, 2016/17,  
 494 2017/18 and 2018/19 academic years. There is fluctuation in the number of Mondays from year to  
 495 year due to the timing of Easter which influences the half-term break as well, which is typically held  
 496 towards the end of May. From Table 6 it can be seen that the average attendance for Mondays in the  
 497 Summer term of 2018/19 was significantly higher than the previous years. Further, not only was the  
 498 2018/19 attendance data higher, but it was also above the required 96% target and the first time that  
 499 this was the case in four years. The range and median for the data also showed the strength of 2018/19  
 500 attendance data when compared to previous years. The range in 2018/19 was over half that of 2015/16  
 501 indicative of a consistently high Monday attendance throughout the term.

**Table 6.** Comparison of Monday Summer term attendance data for I2

	2015/16	2016/17	2017/18	2018/19
Average Attendance (%)	94.4%	94.5%	94.2%	96.5%
Range	6.2%	5.3%	4.5%	2.8%
Median	94.5%	95.0%	94.1%	96.4%
No. of Mondays	12	11	12	10

502 There were some concerns from school leadership on the “stickiness” of Monday Matters events (where  
 503 having an event every other Monday fosters good attendance on other Mondays and indeed other  
 504 days of the week), and whilst there were spikes in attendance on Monday Matters days, attendance  
 505 during the other Mondays was quite good, as evidenced by the data in Table 6. These findings are  
 506 consistent with other studies that noted that in general, pupils are “creatures of habit” who thrive on  
 507 routine, and are thus likely to sustain good attendance once a routine is established [1,2,4].

508 *5.3. Evaluating the Overall Improvement in School Attendance*

509 The full-year attendance comparison is presented in Table 7. It can be seen that initiatives in the  
 510 Spring and Summer terms of the 2018/19 have contributed to an improvement in the whole school  
 511 attendance for the full academic year. Indeed, WPS achieved the required attendance target of 96% for  
 512 the first time in four years in 2018/19.

**Table 7.** Average Attendance for 2015/16, 2016/17, 2017/18 and 2018/19

Session (J <sub>i</sub> )	2015/16	2016/17	2017/18	2018/19
Mon-AM	93.7%	93.7%	93.8%	95.6%
Tues-AM	94.1%	94.9%	94.6%	96.0%
Wed-AM	95.1%	95.0%	95.3%	96.1%
Thur-AM	94.9%	95.3%	94.8%	96.1%
Fri-AM	94.6%	94.7%	94.1%	95.3%
Mon-PM	94.1%	94.5%	94.1%	96.2%
Tues-PM	94.7%	95.6%	95.1%	96.6%
Wed-PM	95.5%	95.6%	95.8%	96.4%
Thur-PM	95.5%	95.7%	95.4%	96.4%
Fri-PM	94.9%	94.9%	94.5%	95.4%
Average AM	94.5%	94.8%	94.5%	95.8%
Average PM	95.0%	95.3%	95.1%	96.2%
Overall (O)	94.7%	95.0%	94.8%	96.0%

513 The data in Table 7 also reveals the success of the Monday Matters initiative on the full-year attendance  
 514 data. Monday AM and PM sessions have seen the largest increase in attendance, with increases of 1.8  
 515 and 2.1 percentage points respectively. As a result, Mondays no longer have the worst-performing AM  
 516 and PM sessions, and the shift in focus now moves towards Fridays, where the underlying reasons for  
 517 poor attendance may be quite different. Unlike Monday absenteeism, which is influenced to some  
 518 extent by separation anxiety, Friday absenteeism may be more influenced by school withdrawal where  
 519 parents may: 1) want to extend the weekend or start holidays earlier to beat the rush and/or save on  
 520 costs, and 2) sometimes assume that Fridays are typically low-value school days in which limited  
 521 learning takes place and hence pursue other activities outside school [21,23]. Hence, the action plan to  
 522 tackle Friday absenteeism must be geared more towards school withdrawal as opposed to the Monday  
 523 Matters initiative which was focused on tackling both school refusal and school withdrawal.

524  
 525 One argument that parents do make on Friday absence is that their child(ren) have excellent attendance  
 526 on all other sessions and these occasional absences should not impact the child and the school. While it  
 527 is well-documented that all and every absence impacts pupil learning, the question of whether Friday  
 528 sessions have now become the most impactful session to overall absence arose [1,4,21]. In line with  
 529 this, the analysis detailed in Sections 4.2.2 and 5.1.1 was conducted on the 2018/19 dataset. It was clear  
 530 from the results in Table 8 that Friday is now the most impactful day to overall below the required  
 531 attendance, with Fri-AM being the most impactful session. Mon-AM is no longer the most impactful  
 532 session to overall below-average attendance for the first time in the four academic years.

**Table 8.** Identifying Target Sessions - 2018/19,  $T_i = 119$ 

Session ( $J_i$ )	$P(J_i)$	$P(J_i, O)$	$\text{conf}(J_i \rightarrow O)$	mt
Mon-AM	0.500	0.467	0.933	-
Mon-PM	0.425	0.392	0.922	-
Tues-AM	0.483	0.467	0.966	-
Tues-PM	0.350	0.333	0.952	-
Wed-AM	0.508	0.492	0.967	-
Wed-PM	0.425	0.408	0.961	-
Thur-AM	0.458	0.433	0.945	-
Thur-PM	0.433	0.408	0.942	-
Fri-AM	0.575	0.525	0.913	0.050
Fri-PM	0.550	0.500	0.909	0.100
Overall ( $O$ )	0.867	-	-	-

### 533 5.3.1. Persistent Absenteeism

534 The impacts of initiatives I1 and I2 on persistent absenteeism (attendance <90%) were also  
 535 analysed with the results presented in Table 9. Persistent absenteeism at WPS has been significantly  
 536 higher than the national average for at least the last three years, this despite regular and close  
 537 monitoring by the school's leadership team (including governors) and the school's attendance officer.  
 538 However, the level of persistent absenteeism has significantly decreased in 2018/19 and was lower  
 539 than the national average for persistent absenteeism of 8.2%.

**Table 9.** Comparison of Persistent Absenteeism

	2015/16	2016/17	2017/18	2018/19
WPS (% of total)	13.1%	11.3%	12.8%	5.8%
National (% of total)	8.2%	8.3%	8.7%	8.2%

540 This is a significant improvement and consistent with previous studies that sought to tackle the  
 541 problem of chronic (persistent) absenteeism, in particular [1]. Indeed some of the approaches for  
 542 tackling persistent absenteeism discussed in [1] have been leveraged in the development of I1 and I2  
 543 including the concept of making rewards more frequent and meaningful.

### 544 5.4. Statistical Testing of the Improvements in School Attendance

545 Statistical testing was conducted using the approach outlined in Section 4.2.5 and the data  
 546 presented in Table 7. For  $H_0^{(1)}$ , the Kruskal-Wallis test showed that there was no statistical difference in  
 547 attendance,  $H = 2.61$ ,  $p < 0.01$ , hence we accept the null hypothesis, while for  $H_0^{(2)}$ , the Kruskal-Wallis  
 548 test showed that there was a statistical difference in attendance,  $H = 21.46$ ,  $p < 0.01$ , hence we  
 549 reject the null hypothesis. Based on this, we thus accept  $H_0^{(1)}$  and reject  $H_0^{(2)}$  and concluded that the  
 550 initiatives in 2018/19 had an impact (positive) on overall attendance.

## 551 6. Conclusions

552 The mt model, described in Equation (3) and detailed in [16], was adapted to improve school  
 553 attendance at WPS. The algorithm detailed in Section 3.2.4, which included the mt model, was used  
 554 to identify the school session which was most impactful to overall below the required average  
 555 attendance. In line with this, the previous three years' attendance data from WPS was analysed  
 556 and it was found that the Monday AM session was consistently the most impactful session to the  
 557 overall below the required average attendance. Two initiatives were carried out at WPS based on  
 558 approaches in previous studies and the collective wisdom of WPS leadership and staff [1,2,20].  
 559 Initiative I1 provided more frequent and meaningful rewards for full attendance while I2 focussed on

improving Monday attendance through the use of themes that were known to be exciting for the pupils.

Both I1 and I2 resulted in a significant improvement of attendance at WPS, with attendance in 2018/19 being at its highest over the past four academic years. Overall average attendance for the 2018/19 academic year was at the required target of 96%, whilst the combined Spring and Summer term attendance was higher at 96.2%. Monday attendance during the Summer term also improved significantly from an average and range perspective. The average Summer term Monday attendance in 2018/19 was significantly higher than the three previous years at 96.5%, while its range was significantly lower 2.8%, implying that attendance on Mondays was consistently better throughout the term.

Analysis of the 2018/19 data using the mt model has revealed that Monday AM is no longer the most impactful session to overall below the required attendance, instead, it is now Friday AM. The underlying dynamics as to why this is the case may also include a shift away from school refusal and more towards school withdrawal (parental condoned absence) which is underpinned by a variety of reasons including cheaper holidays [20,23]. Addressing this is considered to be part of the future work and is detailed in Section 6.2.

### 6.1. Summary of Theoretical and Practical Implications

#### 6.1.1. Theoretical Implications

The proposed approach, which includes the mt model underpinned by well-grounded theory and concepts in tackling absenteeism as detailed in [1] and [26], provides a novel, simple, yet effective way to tackle the well-known problem of addressing absenteeism in schools. The implementation of two, easy-to-action, initiatives have demonstrated a significant improvement in attendance. This study also contributes to the body of knowledge on MBA, in particular, its use in a wide range of sectors including retail, medical and now education [16,32].

#### 6.1.2. Practical Implications

The proposed algorithm detailed in Section 3.2.4 enables schools to easily identify and tackle issues around pupil attendance. This study considered the impact of sessions on attendance, but this approach could be extended to identify other factors impacting attendance including the impact of subjects or topics being taught and the impact of pupil demographics.

The algorithm can also be used to identify other issues at schools as e.g. the factors impacting pupil progress. These factors (which may include attendance, demographics, attentiveness in class and completion of homework) could be quantified using a simple 1 to 5 ranking scale and analysed using the mt model to identify and rank the impact of these factors on pupil progress. Whilst this may be seen as similar to the work in [17], this approach will add further value by quantifying the impact of each factor to overall progress, as opposed to only ranking their association.

### 6.2. Future Work

Future work has been divided into two parts namely: future work for the school; and future work for the authors.

#### 6.2.1. Future work for the school

The school will continue to use the model to keep attendance above the required target. Both the I1 and I2 initiatives are planned for the 2019/20 academic year. At the same time, the school should consider tackling Friday absenteeism, given that Friday is now their new problematic school

604 day. Given that the dynamics may be slightly different as outlined in Section 5.3, the school should  
 605 explore a new series of initiatives, perhaps entitled “Fun-d-mental” Fridays, where the focus is still  
 606 on fun and excitement but also includes the “mental” aspect which emphasizes the need for pupils  
 607 and parents to treat Friday as an essential learning day. Further, the play on the word “fundamental”  
 608 also emphasizes that Fridays are a key part of overall learning (fundamental to learning) as it usually  
 609 involves a consolidation of the week’s work where the various concepts and pieces of work that pupils  
 610 have learned during the work are brought together to both evaluate pupils’ learning and demonstrate  
 611 (to them) how all the learning fits together. It should be noted that schools already use Fridays in this  
 612 way, for example: “Big Write” or “Cold Write” to consolidate the week’s writing activities as well as  
 613 arithmetic testing to assess pupils’ learning and ability to apply the mathematical concepts learned  
 614 during the week [20].

### 615 6.2.2. Future work for the authors

616 The authors have realised, through this study, that school leaders and staff have predominantly  
 617 been trained in the pedagogical aspects of education and thus do not possess advanced skills in  
 618 analytics. In light of this, the authors will investigate automating the proposed approach and include  
 619 a graphical user interface with customisable analytical fields into a software program so that school  
 620 practitioners can benefit from the use of the model across a variety of school fields (e.g. attendance,  
 621 progress, behaviour, etcetera) without the need to conduct detailed programming and data mining by  
 622 themselves. The authors will also consider rolling out the approach and software program to other  
 623 schools so that the benefits and lessons learned at WPS can be shared and maximised.

624 **Author Contributions:** All authors made significant contributions throughout this piece of research and agreed  
 625 to submit the manuscript in the current form. The first author made major contribution in terms of writing  
 626 and implementing software. All the authors contributed in terms of conceptualisation, writing and revising the  
 627 manuscript.

628 **Funding:** This research received no external funding

629 **Acknowledgments:** The authors would like to thank the leadership and staff of Willen Primary School for  
 630 permitting us to use their data and for their efforts in supporting this study, in particular, Ms Emma Warner  
 631 (attendance officer), Ms Carrie Matthews (headteacher), and Ms Sarah Orr (deputy headteacher).

632 **Conflicts of Interest:** The authors declare no conflict of interest.

## 633 Appendix A. Data Tables

634 Detailed data tables are provided that includes the school roll, population size for this study,  
 635 attendance data, and the number of extracted rules.

### 636 Appendix A.1. School Roll and Study Population Size

637 The school roll and study population size is presented in Table A1. Note that children who join  
 638 the roll during the school year were removed from the study population to prevent skewed data, as  
 639 discussed in Section 4.2.2. Children in the Reception year join the school’s attendance roll once they  
 640 turn 5 years old, which almost always occurs after the start of the school year.

**Table A1.** School Roll, and Study Population Size for 2015/16, 2016/17, 2017/18 and 2018/19

	2015/16	2016/17	2017/18	2018/19
School Roll	395	388	382	366
- Reception pupils	57	57	50	46
- In-year mobility	36	20	35	14
Study Population	302	311	297	306
$T_t$	170	180	179	119

641 *Appendix A.2. Attendance Tables*

642 Table A2 details the number of possible sessions for each year, whilst Table A3 details the actual  
 643 attendance record for each session.

**Table A2.** Possible Sessions for 2015/16, 2016/17, 2017/18 and 2018/19

Session ( $J_i$ )	2015/16	2016/17	2017/18	2018/19
Mon-AM	35	35	34	35
Tues-AM	38	38	38	39
Wed-AM	39	39	39	39
Thur-AM	39	39	39	39
Fri-AM	37	39	38	38
Mon-PM	35	35	34	35
Tues-PM	38	38	38	39
Wed-PM	39	39	39	39
Thur-PM	39	39	38	39
Fri-PM	37	39	38	38
Total AM	188	190	188	190
Total PM	188	190	187	190
Overall Total	376	380	375	380

**Table A3.** Actual Average Attendance for each session for 2015/16, 2016/17, 2017/18 and 2018/19

Session ( $J_i$ )	2015/16	2016/17	2017/18	2018/19
Mon-AM	32.8	32.9	31.9	33.5
Tues-AM	35.8	36.3	35.9	37.4
Wed-AM	37.1	37.2	37.2	37.5
Thur-AM	37.0	37.3	37.0	37.5
Fri-AM	35.0	36.4	35.8	36.2
Mon-PM	32.9	33.1	32.0	33.7
Tues-PM	36.0	36.3	36.1	37.7
Wed-PM	37.2	37.3	37.4	37.6
Thur-PM	37.2	37.3	36.3	37.6
Fri-PM	35.1	37.0	35.9	36.3
Total AM	177.7	180.1	177.8	182.1
Total PM	178.4	181.0	177.7	182.9
Overall Total	356.1	361.1	355.5	365.0

644 *Appendix A.3. Number of Rules Extracted*

645 Table A4 details the number of rules extracted per year using minsup = 0.3, and minconf = 0.3.

**Table A4.** Rule Extracted for 2015/16, 2016/17, 2017/18 and 2018/19, minsup = 0.3, minconf = 0.3

	2015/16	2016/17	2017/18	2018/19
Number of Rules	311	535	309	144
Study Population	302	311	297	306
$T_t$	170	180	179	119

646 **References**

- 647 1. Balfanz, R.; Byrnes, V. Chronic absenteeism: Summarizing what we know from nationally available data.  
 648 *Baltimore: Johns Hopkins University Center for Social Organization of Schools* **2012**, *1*, 1–46.
- 649 2. Cook, P.J.; Dodge, K.A.; Gifford, E.J.; Schulting, A.B. A new program to prevent primary school absenteeism:  
 650 Results of a pilot study in five schools. *Children and Youth Services Review* **2017**, *82*, 262–270.
- 651 3. Havik, T.; Bru, E.; Ertesvåg, S.K. School factors associated with school refusal-and truancy-related reasons  
 652 for school non-attendance. *Social Psychology of Education* **2015**, *18*, 221–240.

- 653 4. DfE. School Attendance and Schools: Statutory Guidance. [https://www.gov.uk/government/](https://www.gov.uk/government/organisations/department-for-education/)  
654 [organisations/department-for-education/](https://www.gov.uk/government/organisations/department-for-education/), 2019. Accessed: 2019-07-13.
- 655 5. Robinson, C.D.; Lee, M.G.; Dearing, E.; Rogers, T. Reducing student absenteeism in the early grades by  
656 targeting parental beliefs. *American Educational Research Journal* **2018**, *55*, 1163–1192.
- 657 6. Sukhija, K.; Jindal, M.; Aggarwal, N. The recent state of educational data mining: A survey and future  
658 visions. 2015 IEEE 3rd International Conference on MOOCs, Innovation and Technology in Education  
659 (MITE). IEEE, 2015, pp. 354–359.
- 660 7. Daud, A.; Aljohani, N.R.; Abbasi, R.A.; Lytras, M.D.; Abbas, F.; Alowibdi, J.S. Predicting student  
661 performance using advanced learning analytics. Proceedings of the 26th international conference on world  
662 wide web companion. International World Wide Web Conferences Steering Committee, 2017, pp. 415–421.
- 663 8. Xing, W.; Guo, R.; Petakovic, E.; Goggins, S. Participation-based student final performance prediction  
664 model through interpretable Genetic Programming: Integrating learning analytics, educational data mining  
665 and theory. *Computers in Human Behavior* **2015**, *47*, 168–181.
- 666 9. Qu, S.; Li, K.; Wu, B.; Zhang, S.; Wang, Y. Predicting Student Achievement Based on Temporal Learning  
667 Behavior in MOOCs. *Applied Sciences* **2019**, *9*, 5539.
- 668 10. López-Zambrano, J.; Lara, J.A.; Romero, C. Towards Portability of Models for Predicting Students' Final  
669 Performance in University Courses Starting from Moodle Logs. *Applied Sciences* **2020**, *10*, 354.
- 670 11. Agasisti, T.; Bowers, A.J. 9. Data analytics and decision making in education: Towards the educational data  
671 Scientist as a key actor in schools and higher education institutions. In *Handbook of contemporary education*  
672 *economics*; Edward Elgar Publishing, 2017; p. 184.
- 673 12. Helal, S.; Li, J.; Liu, L.; Ebrahimie, E.; Dawson, S.; Murray, D.J.; Long, Q. Predicting academic performance  
674 by considering student heterogeneity. *Knowledge-Based Systems* **2018**, *161*, 134–146.
- 675 13. Weng, C.H. Mining fuzzy specific rare itemsets for education data. *Knowledge-Based Systems* **2011**,  
676 *24*, 697–708.
- 677 14. Daniel, B. Big Data and analytics in higher education: Opportunities and challenges. *British journal of*  
678 *educational technology* **2015**, *46*, 904–920.
- 679 15. Merceron, A.; Yacef, K.; Romero, C.; Ventura, S.; Pechenizkiy, M. Measuring correlation of strong symmetric  
680 association rules in educational data. *Handbook of educational data mining* **2010**, pp. 245–256.
- 681 16. Moodley, R.; Chiclana, F.; Caraffini, F.; Carter, J. A product-centric data mining algorithm for targeted  
682 promotions. *Journal of Retailing and Consumer Services* **2019**, p. 101940.
- 683 17. Meghji, A.F.; Mahoto, N.A.; Unar, M.A.; Shaikh, M.A. Analysis of Student Performance using EDM  
684 Methods. 2018 5th International Multi-Topic ICT Conference (IMTIC). IEEE, 2018, pp. 1–7.
- 685 18. Dutt, A.; Aghabozrgi, S.; Ismail, M.A.B.; Mahroiean, H. Clustering algorithms applied in educational data  
686 mining. *International Journal of Information and Electronics Engineering* **2015**, *5*, 112.
- 687 19. Klingler, S.; Käser, T.; Solenthaler, B.; Gross, M. Temporally Coherent Clustering of Student Data.  
688 *International Educational Data Mining Society* **2016**.
- 689 20. WPS. Personal Communication with WPS Leadership. personal communication, 2019. Date: 2017-07-10.
- 690 21. Heyne, D.; Gren-Landell, M.; Melvin, G.; Gentle-Genitty, C. Differentiation between school attendance  
691 problems: Why and how? *Cognitive and Behavioral Practice* **2019**, *26*, 8–34.
- 692 22. MK-Council. Guidance for Improving School Attendance. [https://www.milton-keynes.gov.uk/assets/](https://www.milton-keynes.gov.uk/assets/attach/51484/)  
693 [attach/51484/](https://www.milton-keynes.gov.uk/assets/attach/51484/), 2019. Accessed: 2019-07-13.
- 694 23. Money-Saving-Expert. Can you take kids on term-time holidays without being fined? [https://www.](https://www.moneysavingexpert.com/family/school-holiday-fines/)  
695 [moneysavingexpert.com/family/school-holiday-fines/](https://www.moneysavingexpert.com/family/school-holiday-fines/), 2019. Accessed: 2019-03-25.
- 696 24. Hanna, G.L.; Fischer, D.J.; Fluent, T.E. Separation anxiety disorder and school refusal in children and  
697 adolescents. *Pediatrics in Review* **2006**, *27*, 56.
- 698 25. Brown, W.J.; Wilkerson, A.K.; Boyd, S.J.; Dewey, D.; Mesa, F.; Bunnell, B.E. A review of sleep disturbance  
699 in children and adolescents with anxiety. *Journal of sleep research* **2018**, *27*, e12635.
- 700 26. Kearney, C.A.; Graczyk, P. A response to intervention model to promote school attendance and decrease  
701 school absenteeism. *Child & Youth Care Forum* **2014**, *43*, 1–25.
- 702 27. Sibieta, L.; Belfield, C.; others. 2018 annual report on education spending in England, 2018.
- 703 28. Agrawal, R.; Imieliundefinedski, T.; Swami, A. Mining Association Rules between Sets of Items in  
704 Large Databases. Proceedings of the 1993 ACM SIGMOD International Conference on Management

- 705 of Data; Association for Computing Machinery: New York, NY, USA, 1993; SIGMOD '93, p. 207–216.  
706 doi:10.1145/170035.170072.
- 707 29. Chiclana, F.; Kumar, R.; Mittal, M.; Khari, M.; Chatterjee, J.M.; Baik, S.W.; others. ARM–AMO: an efficient  
708 association rule mining algorithm based on animal migration optimization. *Knowledge-Based Systems* **2018**,  
709 *154*, 68–80.
- 710 30. Moodley, R.; Chiclana, F.; Caraffini, F.; Carter, J. Application of uninorms to market basket analysis.  
711 *International Journal of Intelligent Systems* **2018**.
- 712 31. Coghlan, D. *Doing action research in your own organization*; SAGE Publications Limited, 2019.
- 713 32. Tan, P.N. *Introduction to data mining*; Pearson Education India, 2018.
- 714 33. WPS. School Website. <https://willenprimary.com/>, 2019. Accessed: 2017-07-10.
- 715 34. Ofsted. Willen Primary School Inspection Reports. <https://reports.ofsted.gov.uk/provider/21/110388/>,  
716 2019. Accessed: 2017-07-10.

717 © 2020 by the authors. Submitted to *Appl. Sci.* for possible open access publication under the terms and conditions  
718 of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).