# Using AutoML to Analyze the Effect of Attendance and Seat Location on University Student Grades


Ac Hýbl[1] and Germán H. Alférez[2]

[1] matoush@southern.edu
[2] harveya@southern.edu
School of Computing, Southern Adventist University, PO Box 370, Collegedale TN 37315-0370, USA



**Abstract.** A common claim is that class attendance and sitting at the front of a classroom may improve student grades. This study employs Automated Machine Learning (AutoML) to analyze this claim. The data used in this study came from an attendance-tracking system from a private university in Tennessee, USA. The correlation analysis in Microsoft Azure's Machine Learning workspace was performed by training regression models. No correlation was found between student attendance and seat choice and final course grades. The K-means clustering algorithm was used to train clustering models in Microsoft Azure. At $k = 2$ clusters, a cluster with perfect attendance shows a higher average grade than a cluster with a late attendance average. Seat choice within the classroom does not prove important to the clustering models.



**Keywords:** Microsoft Azure, Automated Machine Learning, Regression, K-means clustering, Class attendance, Seat selection


## 1 Introduction

Although recent studies explore the effects of student seating location on social interaction or course engagement [1, 2], there is not yet consensus on the claim that seating location directly affects academic performance [3–6]. Moreover, although there are several research projects that have tried to analyze the relation between seating location and student academic performance, they are focused on small datasets [3, 5, 7, 8].

Our contribution is to apply two machine learning (ML) techniques, namely regression and clustering, to the attendance and seating data from Southern Adventist University (SAU), a private institution located in Tennessee, USA, to analyze the common claim that class attendance and sitting at the front of a classroom may improve student grades. The goal is not only to test whether these algorithms would arrive at conclusions that would support the aforementioned claims, but also to introduce machine learning as a method of analysis for student attendance and seating location. In our study, 221,600 attendance records from $n = 2,067$ students were analyzed from data collected in 2021 and 2022.

To perform machine learning experiments, we made use of the Automated Machine Learning (AutoML) functionality of Microsoft Azure. AutoML was used because of its automated training and evaluation of machine learning models without the need for extensive coding.

This research work answers the following questions:

1. Do class attendance and punctuality foreshadow higher course grades?
2. Do students that sit in the front of class receive higher marks than those that choose to sit near the back?

This paper is organized as follows. Section 2 presents the state of the art. Section 3 presents the methodology. Section 4 presents the results. Section 5 presents the conclusions and future work.

## 2 State of the Art

Regular attendance is widely believed to have numerous benefits for students, including improved learning opportunities and better chances for success. The National Center for Education Statistics, for instance, explains that students who are frequently absent during their schooling years, starting from kindergarten to high school, miss out on crucial learning opportunities, thus hampering their chances for success [9]. Not only does this impede a teacher's ability to facilitate students' success, but research has also shown that students who exhibit frequent absenteeism tend to display a "history of negative behaviors" even after they leave school.

Numerous studies have explored the relationship between student attendance and academic performance. In their 2015 study of first-year psychology courses, Alexander and Hicks analyzed whether class attendance was linked to increased student performance in modern classrooms with online lectures [10]. Their results featured significant ($p < 0.001$ and $p < 0.05$) correlations between student attendance and performance on assignments.

Furthermore, several studies have been done on seat location and student grades. For instance, in a 1973 issue of *Sociometry*, Becker et al. demonstrate that students sitting nearer to their instructor not only received higher grades than those further away, but also liked their professor more ($p < 0.01, n = 282$) [8].

Consistent findings have emerged in a 2017 study by Shernoff et. al. [4]. This study considered whether students ($n = 407$) with particular personality traits chose their seating location preferentially, whether seating location affected subjective experiences such as engagement and attention, and whether these factors affected student performance. To differentiate "causal mechanism[s] from self-selection," the Experience Sampling Method (ESM) method was employed. ESM consists of analyzing a participant's own experience of their surroundings over the course of the study to eliminate the effects of self-selection. Results showed that students in the front rows were more engaged in class and received higher grades than those in the back rows.

Another recent study by Lyu, Jiang, and Wu examined, among several other research questions, the effects of seating on academic achievement in a college setting using $n = 306$ students [3]. This study employed seating networks and clusters and identified the top and bottom performers in each class. Results showed that students with high academic achievement preferred to sit in the front of the classroom while the opposite was also true. One very interesting finding was that while the lowest three performers in each class moved around the classroom over the course of a semester, the top three performers generally remained in the same seats located in the front and middle of a classroom's center columns. In fact, the top of the class *never* moved.

Similarly, Bergtold, Yeager, and Griffin investigated the role of seating location in student performance using a variety of variables [11]. These included gender, grades, math level, GPA, class rank, major, spacial locale, endogenous peer effects, and the effects a student's classmates. The relevant results of this study again supported sitting in the back of the classroom decreased a student's grades. Various statistical analyses have similarly concluded that seating location can impact student academic performance [12].

However, not all studies agree with this body of research. While Chan et al. have generally found that seat location can affect student performance, they differentiate the strength of this correlation by discipline [13]. They found that while students in soft fields such as psychology and sociology tend to perform better when seated in the front of the classroom, seating location has little effect on students in hard fields such as engineering and math.

Two 2007 studies found similar results that student seating did not affect performance [5, 6]. Kalinowski et al. conducted a randomized blind study in a sophomore biology classroom ($n = 43$) using exam scores to assess whether placing students in the front, center, or back of the classroom would affect their grades [5]. They found no evidence that grades or student attitudes were affected by seat location. Moreover, when given the choice, students with high GPAs were found to choose seats in the front of the classroom, suggesting that self-selection may play a role in some of these studies.

In the second study, Armstrong et al. included $n = 5814$ students [6]. Some classrooms used randomly-assigned seating, while others allowed students to choose their own seats and self-report them. Out of 20 classrooms in the study, only seven showed a significant correlation between seat location and student performance. One classroom found that students in the back of the classroom performed better than those in the front. The remaining six allowed for student-chosen seating, thus allowing for self-selection effects. Moreover, the correlations for these classrooms were negligible ($r < 0.16$).

In more recent research (2016), Pichierri and Guido Pichierri and Guido analyzed the effect of seat location on student academic performance and how shyness may be a moderating factor [7]. Data was collected over five years for a total of $n = 232$ regularly-attending students. In general students in the front of the classroom performed better than those in the back. However, this effect decreased as shyness increased.

Despite the above research in this area, we were unable to find studies that incorporated student seat location and machine learning. The most related idea found incorporated simulating student-teacher proximity using an agent-based modeling approach [14]. Moreover, the studies presented in this section tend to be performed on a small scale. Using, few classes, few students, or only one semester of data. Our research uses a larger dataset.

## 3 Methodology

The first step to performing successful experiments in data science is obtaining good data. Therefore, this section first describes where data was obtained, how it was organized, and what precautions were practiced to avoid data-related problems. Next the tools used are introduced along with the procedures and experiments performed.

### 3.1 Data Collection

The data used in the experiments came from an attendance system that records attendance and seating at SAU. Like many other institutions, SAU implemented strict quarantines and digital contact tracing (DCT) in 2020 to enable students to attend classes in person during the COVID-19 pandemic [15]. Thousands of students at SAU were required to use quick response (QR) codes to record their class attendance. The web interface also required students to select their seat within their classroom allowing for digital contact tracing (DCT) whenever a new case of COVID-19 was identified.

As campus activity returned to normal and concern over COVID-19 subsided over the next few semesters, the new attendance system remained. Many professors simply found the system much more convenient than manually taking note of absent or late students. Not only did the new system allow professors to discuss course material sooner, but it continued collecting valuable attendance and seating data which proved to be useful for more than just contact tracing.

Because the driving questions of this research concern a typical classroom setting, the data collected needed to reflect only this setting. Some courses at SAU only had a small number of students per semester. Thus, only large classrooms were considered. Additionally, because stickers on the desks in each classroom indicated the seat row and column (as shown in Figure 1), only classrooms with bolted desks were selected. This ensured that rearranged classrooms did not introduce flawed data into the study.

Some faculty raised further concerns that the attendance system had changed as the university relaxed its COVID-19 restrictions. For instance, during its first semester of use, the attendance tracking system (ATS) only permitted students to sit in every other seat, thus restricting student seat choice. However, in following semesters, ATS allowed students to sit in any seat in a classroom. In favor of consistency, only the latter of these systems was used, resulting in two semesters of data (Fall 2021 and Winter 2022).

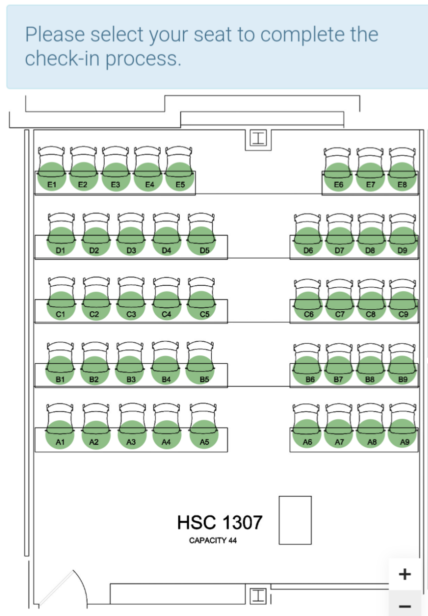


Fig. 1: The ATS interface in one of the many classrooms on campus.

After selecting locations and times for data collection, all other entities were determined as shown in Figure 2. 159 course sections with twenty or more students enrolled were found using the seventeen chosen classrooms over the two semesters. 2,067 students were enrolled in one or more of these sections, which were taught by sixty-three professors representing thirteen departments. The final attendance dataset not only represented most of the diversity at the university, but also consisted of nearly a quarter million (221,600) rows. Therefore, this data-collection method improves upon previous studies with fewer data points [3, 4, 8].

Each section had associated events that represented one class period. The ATS stored data for each student that was present, but it did not always specify if a student was absent from a class. Thus, using several structured query language (SQL) scripts, this data was imputed. If a student was enrolled in a section but had no record of attending any one of its events, they were assumed absent.

### 3.2 Data Organization and Tools

Dividing classes, students, and attendance into separate relations was the most natural way of storing and organizing the data. Because the data was pulled from a data warehouse partially external to the university, it was not separated

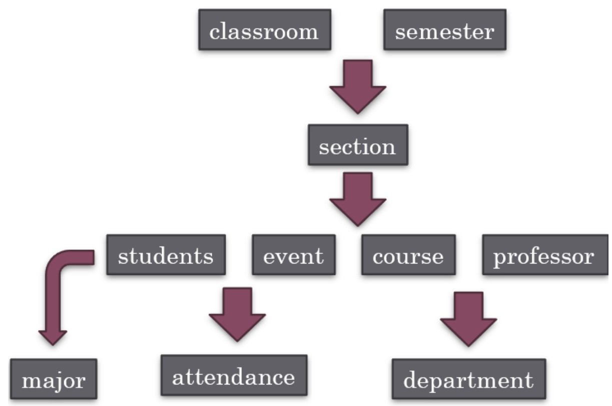


Fig. 2: The natural organization of entities in data collection. Choosing classrooms and semesters resulted in specific sections. These sections had specific professors, students, and events. Attendance is a many-to-many relationship between students and events.

this way. To structure the data, several segregation and cleaning steps were performed in Tableau Prep Builder.[3]

Because Microsoft Azure was to be used for the data science experiments, it was also chosen to store and serve the data as well. Using an Azure SQL Database on an Azure SQL Server, we constructed a relational database from the outputs of Tableau Prep Builder. After resolving all the bugs encountered during data collection, the cleaned comma-separated values (CSV) files provided by Tableau Prep Builder were simply imported as tables into the database using Microsoft SQL Server Management Studio (SSMS). Primary and foreign keys were also configured in SSMS. The resulting schema for this database can be seen in Figure 3. All entities are transitively related using primary/foreign keys.

The initial reasons for using Microsoft Azure were its advertized ease of use and AutoML workspace. AutoML is an emerging technology that offers automatic training of various machine learning models without the need for coding. In a typical solution, data is first supplied to a model training component. The component trains a prediction model, and may even automatically select the most appropriate algorithm for solving a given classification, regression, clustering, or forecasting problem. It may also tune the model's hyperparameters.

All experiments were run on a Microsoft Azure's cloud computing platform. The specific machine configuration selected, Standard D2 v2, had the following specifications: Cores: 2; RAM: 7 GB; disk: 100 GB; temporal storage (SSD): 100 GiB; NICs: 2; network bandwidth: 1500 Mbps; throughput IOPS: 8x500; maximum data disks: 8; and maximum temporal storage throughput: IOPS/Read MBps/Write MBps: 6000/93/46.

[3] tableau.com/products/prep

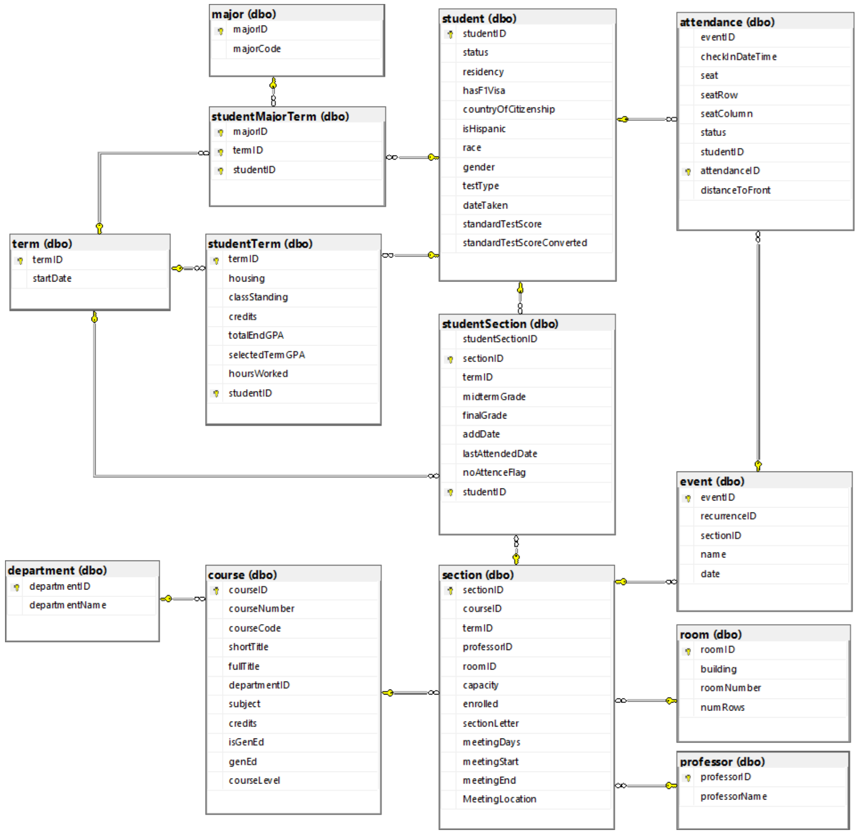


Fig. 3: The database schema as visualized by Microsoft SQL Server Management Studio.

### 3.3 Experiment Plan

The general plan for regression and clustering was to connect the Microsoft Azure AutoML workspace to the Microsoft Azure SQL Database so that the AutoML components could automatically extract the latest version of the data from the database. Using these two platforms, we created automatic pipelines that built machine learning models for grade forecasting. Finally, various subsets of attributes were provided to the AutoML pipeline to run experiments on.

Once a model was created, Microsoft Azure supplied various metrics associated with that model. For regression, these included error scores and correlation coefficients, allowing for a simple correlation analysis of any subset of attributes. For clustering, metrics included cluster densities and diameters as well as each record's cluster assignment. Using this information, the model could provide the average value of each attribute for each cluster, which could give insight into how the algorithm naturally organized the data.

To perform logistic regression and clustering, class data first needed to be transformed into a numerical format. The twelve grade categories "A-F" were converted to the numbers 1-12, respectively. Also, "I" (incomplete) and "IP" (incomplete passing) were assigned values of 13 and 14.[4]

Other categorical variables were converted in a similar manner. For example, there were five categories for attendance status. The labels "Present," "Online," "Late," "Excused," and "Absent" were assigned the values 0-4 respectively.

Further, as shown in the ATS interface in Figure 1, students selected their seat using a numerical column and a *row letter*. Most training models would perform better with a *row number* rather than a letter. Moreover, the number of rows and spacing between those rows in each classroom varied, rendering any categorical row data inconsistent. To provide the most useful data to the algorithms that would train the models, the row letters were extracted, aggregated for each classroom, and converted to a normalized distance from the front of the classroom. This new attribute, called *distanceToFront*, measured how far a student's chosen seat was from the front of the classroom. Values closer to "0" indicate seats closer to the front row of a classroom while those closer to "1" represent seats at the back of a classroom.

The final query fetches the attributes of interest for this project (shown in Listing 1.1).[5] These attributes included student demographic information, credit load, hours worked during the semester at the university, distance to the front of the classroom for every attendance record, attendance status, and final course grade. This query was run against the Azure SQL database and the resulting data was used as a starting point for Azure's AutoML experiments.

```sql
select s.isFemale, s.isHispanic, s.race, st.housing,
  st.gradeLevel, st.credits, st.hoursWorked,
  a.distanceToFront, a.col, a.statusCode, sn.finalGradeCode
from attendance a join student s on a.studentID = s.
 studentID
  join studentTerm st on s.studentID = st.studentID
  join studentSection sn on s.studentID = sn.studentID
    and sn.termID = st.termID
  join section n on sn.sectionID = n.sectionID
  join event e on n.sectionID = e.sectionID
    and e.eventID = a.eventID
  join course c on n.courseID = c.courseID
where c.department != 'NRSG' and c.department != 'PEAC';
```

[4] This assumes that not completing a class is a less favorable outcome than failing it. Also "Incomplete Passing" is marked as lower than "Incomplete," but it is not a cause for concern as this represents less than 0.1% of the data.

[5] Notice that courses in the Nursing (NRSG) and Physical Education (PEAC) departments were filtered out. Nursing courses were removed because students were often assigned seats in these courses, thus removing the student's ability to choose their seat. Other classrooms allowed students to choose their own seat, and we preferred not to mix these two techniques in this study. Physical education courses, on the other hand, were not considered a "normal classroom setting."

Listing 1.1: Query to fetch the attributes of interest.

**Correlation Analysis** Regression experiments were configured as "jobs" and started in Microsoft Azure's Machine Learning workspace. To analyze correlation across different groups of attributes, the experiment was run multiple times with different subsets of the columns shown in the query shown in Listing 1.1. The platform automatically selected the best-preforming algorithms based on grade-prediction power.

**Clustering** Rather than telling the algorithm how the data should be fit, unsupervised learning groups data based on all attributes without prior direction. Clusters formed by minimizing a cost function can be assessed for conclusions.

Microsoft Azure's AutoML platform can perform popular clustering algorithms such as K-means. The best way to use clustering in Microsoft Azure is to create a pipeline. In a pipeline, several components can be "wired together" to create a single process. This process is capable of gathering data, processing it, training machine learning models, testing those models, and generating result data in automatic succession. The machine learning pipeline shown in Figure 4 was used to perform clustering for this project. Note that to make this image fit better on paper, a data cleaning step and a column selection step were removed from the pipeline. The cleaning step simply imputed missing values using column means.

The **Import Data** component fetches dynamic data from the Azure SQL Database. Different attributes can be selected within or after the query that this component performs. Initially, numerical, textual, and categorical data was being returned. However, after several experiments and database adjustments, only numerical data was fetched using the query shown in Listing 1.1.

On the top left of Figure 4, the **K-Means Clustering** component is used to initialize and configure an untrained clustering model. Several parameters can be set here: the number of clusters desired in the output, the feature normalization option, a model weight initialization algorithm, and a multi-dimensional distance metric.[6] The **Train Clustering Model** component then trains this model using the imported data.

The **Assign Data to Clusters** component takes all the provided data and assigns it to a cluster using the trained model. Using this data, the model can be evaluated in the **Evaluate Model** component, which measures average distances between all clusters.

Microsoft Azure provides a mechanism for executing Python code in a pipeline. A custom Python script only needs to contain an `azureml_main()` function that

[6] In our experiments, the number of clusters varied from two to fourteen, normalization was enabled, weights were initialized with the "K-Means++" algorithm, and a Euclidean distance metric was used.

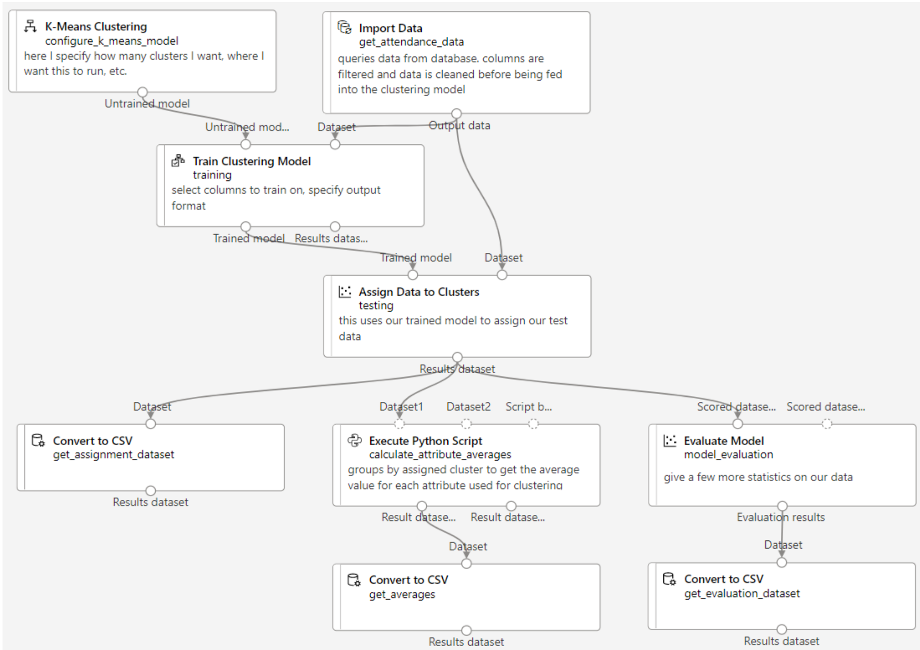


Fig. 4: The clustering pipeline created in this project.

receives and returns up to two dataframes. To generate the necessary data, the **Execute Python Script** component receives all the records along with their cluster assignments and calculates the average value of every feature for each cluster (see Listing 1.2).

```python
import pandas as pd
def azureml_main(dataset, optional_data = None):
    return dataset.groupby("Assignments").mean()
```

Listing 1.2: The data was grouped by cluster and then aggregated into a mean.

Finally, all the data generated by these components is converted into CSV format as described in the following section.

# 4 Results

This section presents the results of the correlation and clustering analyses.

## 4.1 Correlation Analysis

In the first regression experiment, all of the attributes of interest mentioned in the previous section were fed into the regression model.[7] The Root Mean

[7] This table summarizes the the findings of four regression experiments. The first included demographics, course and work amounts, attendance information, and *final-*

Squared Error (RMSE), Explained Variance (EV), Spearman Correlation Coefficient, and $R^2$ Score are presented in Table 1. This configuration provided a fairly accurate regression model. The Spearman correlation coefficient was over 0.9, suggesting a high correlation between all the input and target attributes. As a visual representation of its accuracy, the regression graph is shown in Figure 5.

Table 1: Regression Performance Metrics

| Attributes Used | RMSE | EV | Spearman | $R^2$ Score |
|---|---|---|---|---|
| All shown in SQL query | 1.110 | 0.859 | 0.906 | 0.859 |
| statusCode, distanceToFront | 2.314 | 0.043 | 0.186 | 0.043 |
| statusCode | 2.923 | 0.020 | 0.107 | 0.020 |
| distanceToFront | 2.922 | 0.020 | 0.140 | 0.020 |

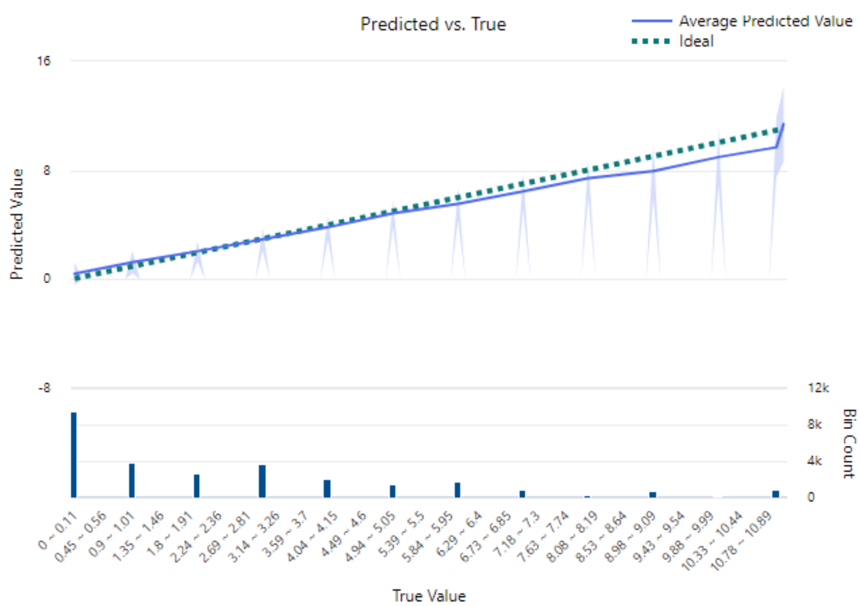


Fig. 5: The predicted values approximate the true values very well.

Microsoft Azure also provides insight into which attributes were most important to the accuracy of the model. As shown in Figure 6, the model relied

*GradeCode*, which is the target variable. The next three experiments only included distance to the front of the classroom or attendance status, or both.

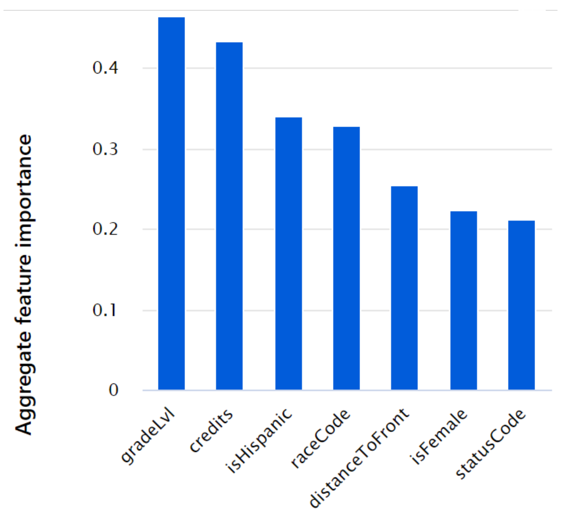


Fig. 6: The top seven attributes used by the first regression model, ranked in order of predictive importance.

heavily on class standing, term information, and demographic information. The distance to the front of the classroom falls in fifth place and attendance status in seventh.

As evidenced, the predictive capacity of a student's ultimate grade can be derived with a noteworthy level of precision by examining their grade level, workload, and demographics. Given the model's underlying objective of optimizing accuracy, repeated training iterations would fail to encourage greater weighting of the ($distanceToFront$) and ($statusCode$) features. Consequently, it was necessary to limit the model's scope to only these two attributes. The following experiment used only attendance status and distance to the front of the classroom. Microsoft Azure took forty-eight minutes to converge on an optimal model. This model offered a mean absolute error that was nearly a fifth of the entire range. Very little of this error was explained by the variance in attendance data. Additionally, both correlation metrics suggested no correlation between the attendance data and student grades. In fact, the $R^2$ score suggests that there is a 95.7% chance of getting this correlation from unrelated attributes.

Isolating the attendance status and seat row attributes even further did not improve the results. Both prediction models produced separately by these two attributes yielded even lower correlation values.

Overall, the results of the correlation analysis showed that attendance and seat choice could not be used to accurately or precisely predict student grades in the data obtained from SAU.

### 4.2 Clustering

Before other hyperparameters were tuned, the number of clusters, $k$, had to be decided. To this end, we made use of the Elbow Method. As shown in Figure 7, the inflection point is slightly unclear. Thus, experiments were first performed with $k = 5$ clusters. The first experiment used all the queried columns. However, the clustering algorithm grouped data primarily using demographic information. For example, four of the five clusters would be entirely based on a student's gender and Hispanic status.

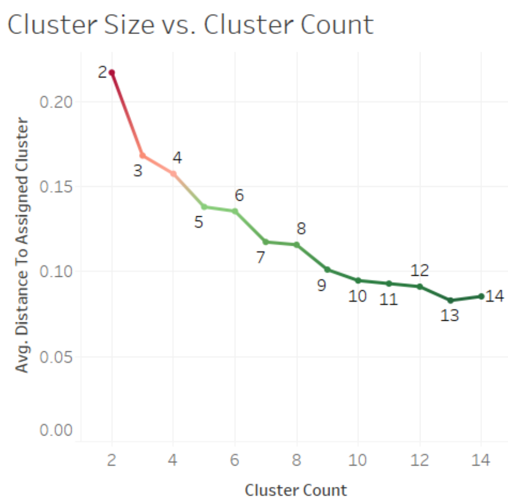


Fig. 7: The average distance to a cluster compared across various number of clusters.

After limiting the number of attributes available, K-Means grouped the data primarily based on attendance status and distance to the front of the classroom. However, five clusters proved too many for this set of attributes so the count was reduced to $k = 3$.

The averages in Table 2[8] were obtained using three clusters and two input attributes. Data placed in Cluster 0 represents a mostly "Present" attendance status and a seat location roughly halfway from the front of the classroom. Nearly all of the attendance records in this cluster have perfect scores of "A". Cluster 1 has similar averages, but with grades much closer to "C+" and "C". The cluster with chronically late attendance and a seating preference slightly beyond the halfway point has an average grade around a "B+" or "B". Although Cluster 2

[8] The radius feature was originally named "Average Distance to Cluster Center" and has been scaled so that the largest value is 1.

does show a group of students that is often late and performs worse than average, the three clusters together are inconclusive.

Table 2: Average Attribute Values in Three Clusters

| | status | distanceToFront | grade | radius | point count |
|---|---|---|---|---|---|
| Cluster 0 | 0.020 | 0.575 | 1.004 | 0.404 | 104,833 |
| Cluster 1 | 0.030 | 0.613 | 6.384 | 0.569 | 42,212 |
| Cluster 2 | 2.922 | 0.610 | 3.618 | 1.000 | 18,143 |
| Total | 0.341 | 0.589 | 2.666 | 0.512 | 165,188 |

Finally, the experiments were repeated with only two clusters. The results are displayed in Table 3. As seen before, distance to the front of the classroom is nearly the same for both clusters. However, Cluster 1 specifically represents records where a student was, on average, late to class. This cluster has an average grade between "B+" and "B". Cluster 0, on the other hand, displays much better attendance and grades on average between "A-" and "B+".

Table 3: Average Attribute Values in Two Clusters

| | status | distanceToFront | grade | radius | point count |
|---|---|---|---|---|---|
| Cluster 0 | 0.022 | 0.586 | 2.548 | 0.689 | 147,033 |
| Cluster 1 | 2.921 | 0.610 | 3.624 | 1.000 | 18,155 |
| Total | 0.341 | 0.589 | 2.666 | 0.651 | 165,188 |

Though interesting, these clusters are still far from ideal. One major issue is that they are unbalanced. One includes 89% of the attendance data and the other only 11%. A more balanced dataset is desirable.

To further validate Microsoft Azure's automated clustering algorithm, clustering was also performed in Weka 3.[9] The results from this experiment were identical to those obtained from Microsoft Azure's K-Means clustering.

### 4.3 Discussion

The correlation analysis did not support any correlation between a student's attendance and their performance in class ($R^2 = 0.020$). Clustering provided more insight. A cluster of data was identified that represented tardiness or absence along with lower grades. No experiments showed that sitting nearer to the front of the classroom positively impacted a student's grades. Instead, this attribute seemed generally unimportant.

[9] cs.waikato.ac.nz/ml/weka

## 5 Conclusion and Future Work

In this study, after applying regression with AutoML, no correlation was found between student attendance and seat choice and final course grades. Also, the K-means analysis shows two clusters, one of them with perfect attendance and a higher average grade than the other with a late attendance average. This reveals that unsupervised learning through clustering does support the claim that attendance has a positive effect on student performance based on the data used in this study. However, seat choice within the classroom did not prove important even in the unsupervised learning analysis.

As future work, we will explore other technologies for machine learning because Microsoft Azure's AutoML proved difficult to navigate and configure. Also, we plan the run more experiments with balanced data to include the similar amounts of high and low scores. Specifically, more data with final grades below "B" should be collected and combined with the current dataset. Also, a better balance is needed between the different attendance categories. Currently, "Late," "Absent," and "Excused" attendance records represent only 11.6% of attendance data.

Also, we plan to evaluate additional variables in the study. For instance, the relative distance of students to the front of smaller and larger classrooms should be considered. Also, data about the horizontal placement of a student within a classroom will be included as well as teaching styles, tardiness thresholds, and class time.

## References

1. J. Wang, K. Xie, Q. Liu, T. Long, and G. Lu, "Examining the effect of seat location on students' real-time social interactions in a smart classroom using experience sampling method," *Journal of Computers in Education*, pp. 1–19, 2022.
2. N. Gao, M. S. Rahaman, W. Shao, K. Ji, and F. D. Salim, "Individual and group-wise classroom seating experience: Effects on student engagement in different courses," *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, vol. 6, no. 3, sep 2022. [Online]. Available: https://doi.org/10.1145/3550335
3. Q. Lyu, Y. Jiang, and J. Wu, "Relations between university students' academic achievement and their seating positions in classrooms," in *2021 7th International Conference on Education and Training Technologies*, 2021, pp. 36–43.
4. D. J. Shernoff, A. J. Sannella, R. Y. Schorr, L. Sanchez-Wall, E. A. Ruzek, S. Sinha, and D. M. Bressler, "Separate worlds: The influence of seating location on student engagement, classroom experience, and performance in the large university lecture hall," *Journal of Environmental Psychology*, vol. 49, pp. 55–64, 2017.
5. S. Kalinowski and M. L. Toper, "The effect of seat location on exam grades and student perceptions in an introductory biology class." *Journal of College Science Teaching*, vol. 36, no. 4, 2007.
6. N. Armstrong and S.-M. Chang, "Location, location, location." *Journal of College Science Teaching*, vol. 37, no. 2, 2007.
7. M. Pichierri and G. Guido, "When the row predicts the grade: Differences in marketing students' performance as a function of seating location," *Learning and Individual Differences*, vol. 49, pp. 437–441, 2016.

8. F. D. Becker, R. Sommer, J. Bee, and B. Oxley, "College classroom ecology," *Sociometry*, pp. 514–525, 1973.
9. NCES, "Every school day counts: The forum guide to collecting and using attendance data," Feb 2009. [Online]. Available: https://nces.ed.gov/pubs2009/attendancedata/chapter1a.asp
10. V. Alexander and R. E. Hicks, "Does class attendance predict academic performance in first year psychology tutorials?" *International Journal of Psychological Studies*, vol. 8, no. 1, 2016.
11. J. S. Bergtold, E. A. Yeager, and T. W. Griffin, "Spatial dynamics in the classroom: Does seating choice matter?" *PLOS ONE*, vol. 14, no. 12, pp. 1–16, 12 2020. [Online]. Available: https://doi.org/10.1371/journal.pone.0226953
12. B. J. Minchen, "The effects of classroom seating on students' performance in a high school science setting," Ph.D. dissertation, 2007.
13. K. L. Chan, D. C. Chin, M. S. Wong, R. Kam, B. S. B. Chan, C.-H. Liu, F. K. K. Wong, L. K. Suen, L. Yang, S. C. Lam *et al.*, "Academic discipline as a moderating variable between seating location and academic performance: implications for teaching," *Higher Education Research & Development*, vol. 41, no. 5, pp. 1436–1450, 2022.
14. Z. Dong, H. Liu, and X. Zheng, "The influence of teacher-student proximity, teacher feedback, and near-seated peer groups on classroom engagement: An agent-based modeling approach," *Plos one*, vol. 16, no. 1, p. e0244935, 2021.
15. M. L. Malloy, L. Hartung, S. Wangen, and S. Banerjee, "Network-side digital contact tracing on a large university campus," in *Proceedings of the 28th Annual International Conference on Mobile Computing And Networking*, ser. MobiCom'22. New York, NY, USA: Association for Computing Machinery, 2022, pp. 367–380. [Online]. Available: https://doi.org/10.1145/3495243.3517029