Data-Driven Course Insights: Predicting Grade Trends

Group 5: Arlette Diaz, Marianne Hernandez, Nandini Jirobe, Sharadruthi Muppidi, Sonina Mut, Yuting Lu

Formulate Question or Problem

Goal

Predict course grade distributions and popularity rankings for upcoming semesters

Problem

Students lacking information when selecting courses

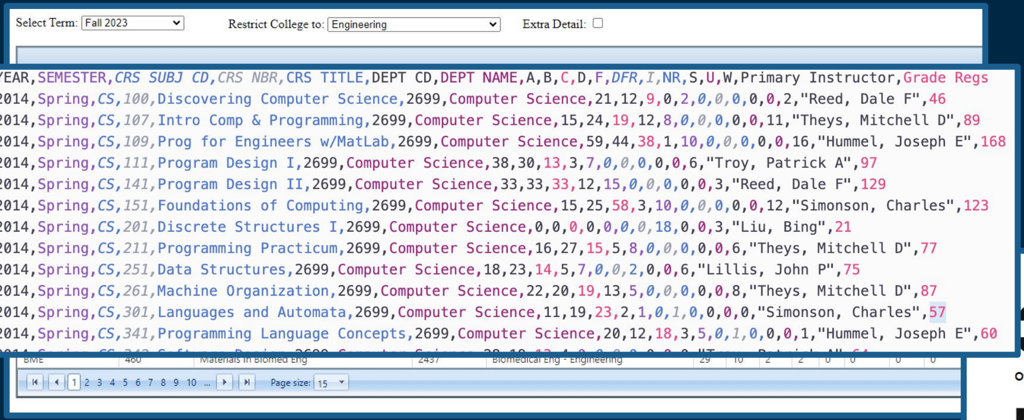
Hypothesis

The historical grade distributions of a class and Rate My Professors ratings can be used to predict future grade distributions and demand for upcoming semesters.

Success Metrics

Provide accurate predictions of course grade distributions and popularity to help students make informed course selection decisions.

Acquire and Clean Data

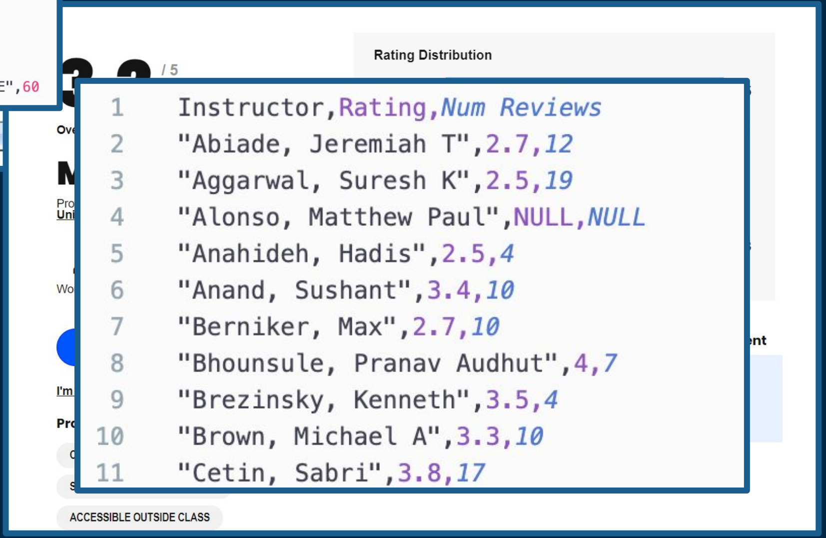

Select Term: Fall 2023 Restrict College to: Engineering Extra Detail:

```
YEAR,SEMESTER,CRS SUBJ CD,CRS NBR,CRS TITLE,DEPT CD,DEPT NAME,A,B,C,D,F,DFR,I,NR,S,U,W,Primary Instructor,Grade Regs
2014,Spring,CS,100,Discovering Computer Science,2699,Computer Science,21,12,9,0,2,0,0,0,0,0,2,"Reed, Dale F",46
2014,Spring,CS,107,Intro Comp & Programming,2699,Computer Science,15,24,19,12,8,0,0,0,0,0,11,"Theys, Mitchell D",89
2014,Spring,CS,109,Prog for Engineers w/MatLab,2699,Computer Science,59,44,38,1,10,0,0,0,0,0,16,"Hummel, Joseph E",168
2014,Spring,CS,111,Program Design I,2699,Computer Science,38,30,13,3,7,0,0,0,0,0,6,"Troy, Patrick A",97
2014,Spring,CS,141,Program Design II,2699,Computer Science,33,33,33,12,15,0,0,0,0,0,3,"Reed, Dale F",129
2014,Spring,CS,151,Foundations of Computing,2699,Computer Science,15,25,58,3,10,0,0,0,0,0,12,"Simonson, Charles",123
2014,Spring,CS,201,Discrete Structures I,2699,Computer Science,0,0,0,0,0,0,0,18,0,0,3,"Liu, Bing",21
2014,Spring,CS,211,Programming Practicum,2699,Computer Science,16,27,15,5,8,0,0,0,0,0,6,"Theys, Mitchell D",77
2014,Spring,CS,251,Data Structures,2699,Computer Science,18,23,14,5,7,0,0,2,0,0,6,"Lillis, John P",75
2014,Spring,CS,261,Machine Organization,2699,Computer Science,22,20,19,13,5,0,0,0,0,0,8,"Theys, Mitchell D",87
2014,Spring,CS,301,Languages and Automata,2699,Computer Science,11,19,23,2,1,0,1,0,0,0,0,"Simonson, Charles",57
2014,Spring,CS,341,Programming Language Concepts,2699,Computer Science,20,12,18,3,5,0,1,0,0,0,1,"Hummel, Joseph E",60
```

Page size: 15

01 UIC Grade Distribution

02 Rate My Professor


Rating Distribution

```
Instructor,Rating,Num Reviews
"Abiade, Jeremiah T",2.7,12
"Aggarwal, Suresh K",2.5,19
"Alonso, Matthew Paul",NULL,NULL
"Anahideh, Hadis",2.5,4
"Anand, Sushant",3.4,10
"Berniker, Max",2.7,10
"Bhounsule, Pranav Audhut",4,7
"Brezinsky, Kenneth",3.5,4
"Brown, Michael A",3.3,10
"Cetin, Sabri",3.8,17
```

ACCESSIBLE OUTSIDE CLASS

Acquire and Clean Data

CS 109

Programming for Engineers with MatLab

3 hours. Credit is not given for CS 109 if the student has credit for CS 111 or CS 112 or CS 113. Extensive computer use required. Prerequisite(s): Credit or

```
Course Code,Course Title,Credits,CRN,Section Type,Time,Days per Week,Instructor,Method,Semester,
CS 100,Discovering Computer Science,3,17397,LCD,02:00 PM – 02:50 PM,3,"Reed, D",On Campus,Spring
CS 107,Introduction to Computing and Programming,4,17412,LEC,12:30 PM – 01:45 PM,2,"Theys, M",On
CS 109,C/C ++ Programming for Engineers with MatLab,3,19466,LCD,02:00 PM – 02:50 PM,2,"Hummel, J
CS 111,Program Design I,3,34013,LCD,02:00 PM – 03:15 PM,2,"Troy, P",On Campus,Spring,2014,Aftern
CS 141,Program Design II,3,34447,LCD,01:00 PM – 01:50 PM,3,"Reed, D",On Campus,Spring,2014,After
CS 151,Mathematical Foundations of Computing,3,34014,LEC,12:00 PM – 12:50 PM,3,"Simonson, C",On
CS 201,Data Structures and Discrete Mathematics I,4,17418,LCD,12:00 PM – 12:50 PM,1,"Liu, B",On
CS 201,Data Structures and Discrete Mathematics I,4,N/A,LCD,12:30 PM – 01:45 PM,2,N/A,On Campus,
CS 211,Programming Practicum,2,34456,LCD,10:00 AM – 10:50 AM,1,"Theys, M",On Campus,Spring,2014,
```

36426	LBD - BAG	10:00 AM - 11:50 AM	F	2249E	2SELE	Riazi, S	Meet on campus
36427	LBD - BAH	12:00 PM - 01:50 PM	F	2249E	2SELE	Riazi, S	Meet on campus

03 Class Scheduler Data

04 Google Scholar

Sathya N. Ravi

FOLLOW

Assistant Professor Of Computer Science at University of Illinois at Chicago

```
Instructor,Interests
"Asudeh, Abolfazl","Responsible Data Science, Algorithmic Fairness, Data Ma
"Bell, John T",
"Bello Lander, Gonzalo Alejandro",
"Burton, Emanuelle Neuman",
"Buy, Ugo A","Software Engineering, Privacy in Social Networks"
"Caragea, Cornelia Alexandra","Natural Language Processing, Deep Learning,
"Chakraborti, Anrin","Computer Security, Cryptography"
"Chattopadhyay, Debaleena","Human-Computer Interaction, Older Adults, Gestu
"Cheng, Lu","Socially Responsible AI, Causal Machine Learning, Data Mining,
```

Proceedings of the AAAI conference on artificial intelligence 33 (01), 4772-4779

Fuzzy assessment of FMEA for rotary switches: a case study | 45 | 2012

S Vinodh, S Aravindraj, SN Ravi, N Yogeshwaran

The TQM Journal 24 (5), 461-475

Exploratory Data Analysis

Data Limitations

- ❖ Incomplete faculty data and course inconsistency
- ❖ Online course data limited to post-COVID
- ❖ Missing grade data for small classes (<5 students)

Key Hypotheses

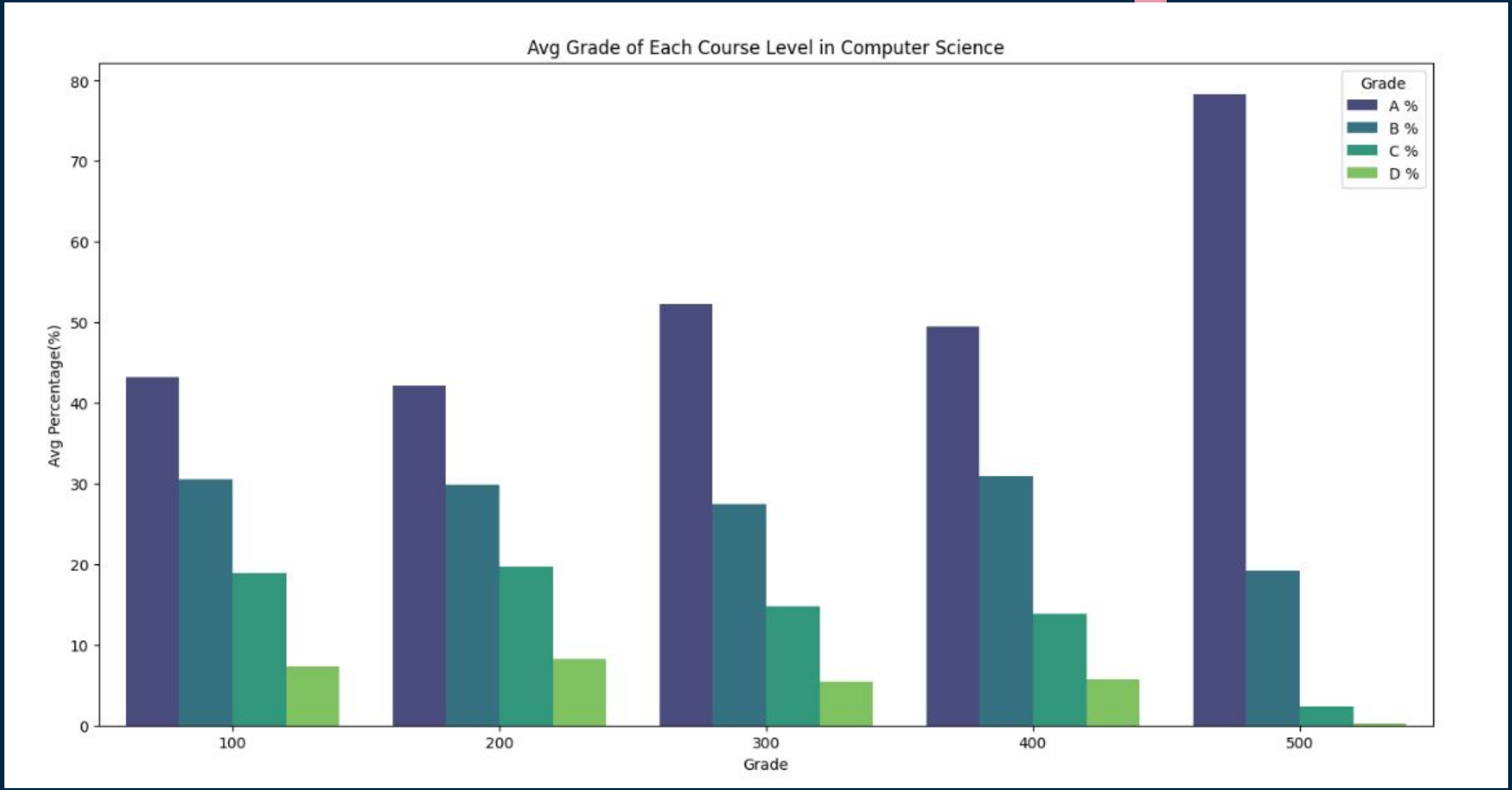
- ❖ RMP reviews show grade-based bias
- ❖ COVID lowered grades (2020)
- ❖ CS has more online offerings than IE/ME
- ❖ Popular professors give higher grades

```
Top 5 Highest Rated CS Professors (minimum 5 reviews):
                              Instructor  Rating  Num Reviews
24  Ganchinho de Pina, Luis Gabriel     5.0         29.0
55                    Medya, Sourav     5.0          7.0
2   Bello Lander, Gonzalo Alejandro     4.9        111.0
53                Maratos, George P     4.9         27.0
14                DasGupta, Bhaskar     4.8         23.0

Top 5 Most Revi

20
73
48
1
2   B
```

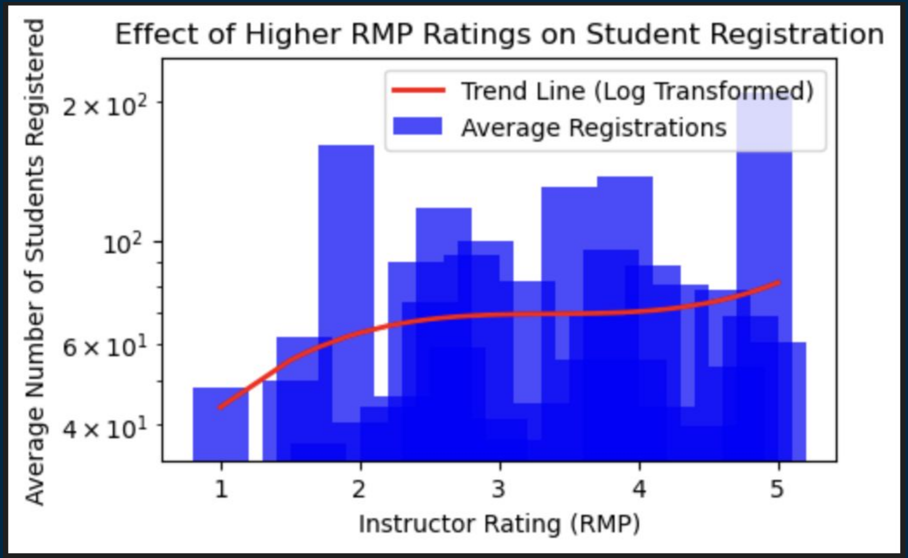
```
Top 5 Courses by GPA (minimum 20 students):
    course_code                CRS TITLE  gpa  total_students
15        CS377  Communication and Ethics  4.0              42
                                           4.0              36
                                           4.0              26
                                           4.0              30
                                           4.0              21

s):
 TITLE       gpa  total_students
licy    1.548387              62
ures    1.688235             170
gn I    1.729730              74
gn I    1.954545              66
n II    1.964103             195
```

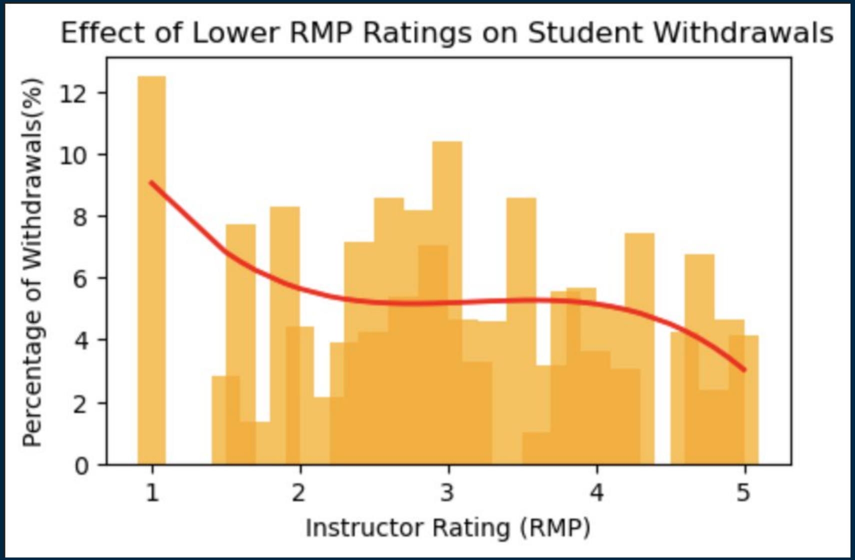
```
Total number of course sections: 2072
Number of unique courses: 110
Number of unique instructors: 153
Year range: 2014 - 2024

Semesters offered:
Semester
Spring    1002
Fall       950
Summer     120
Name: count, dtype: int64
Time of Day
afternoon    1227
morning       669
evening       117
```

Data Visualization # 1: Avg Grade based on Course Level

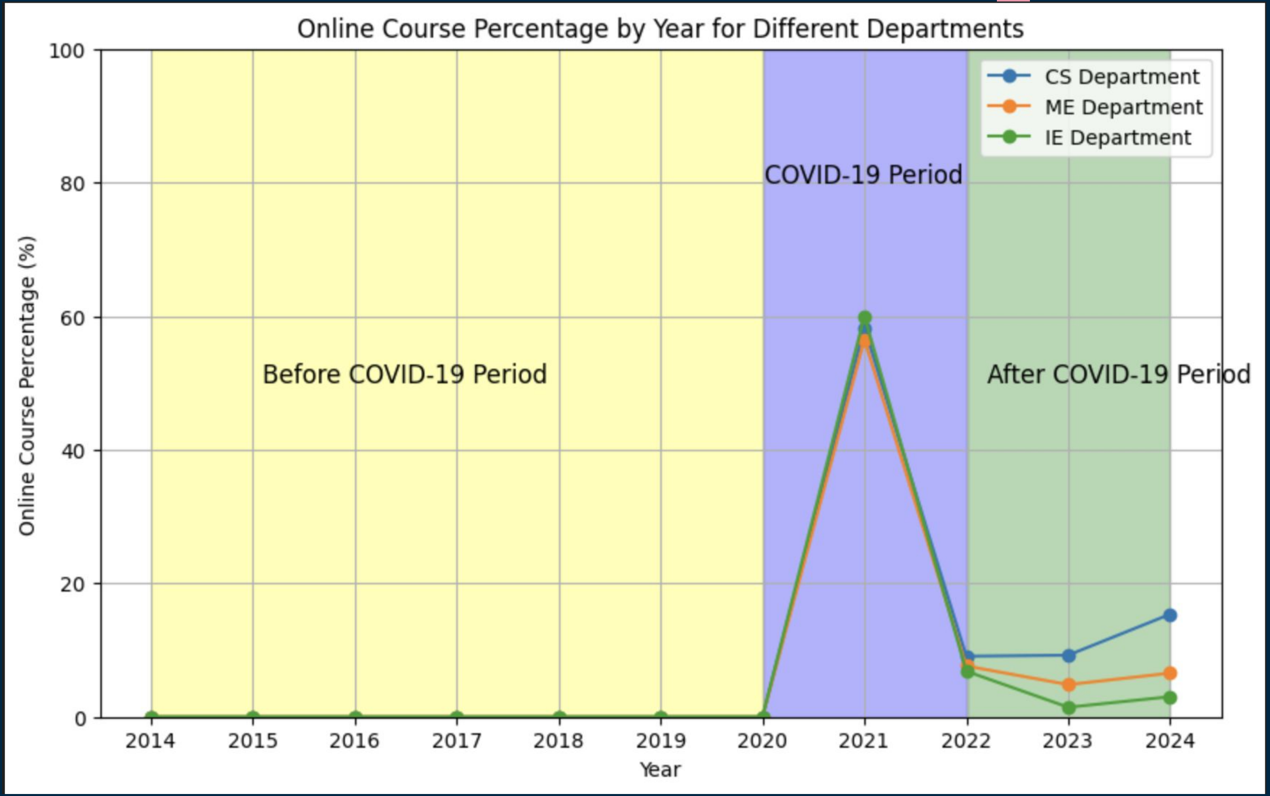


Data Visualization # 2: RMP and Student Retention





Data Visualization # 3: Online Courses per Dept.



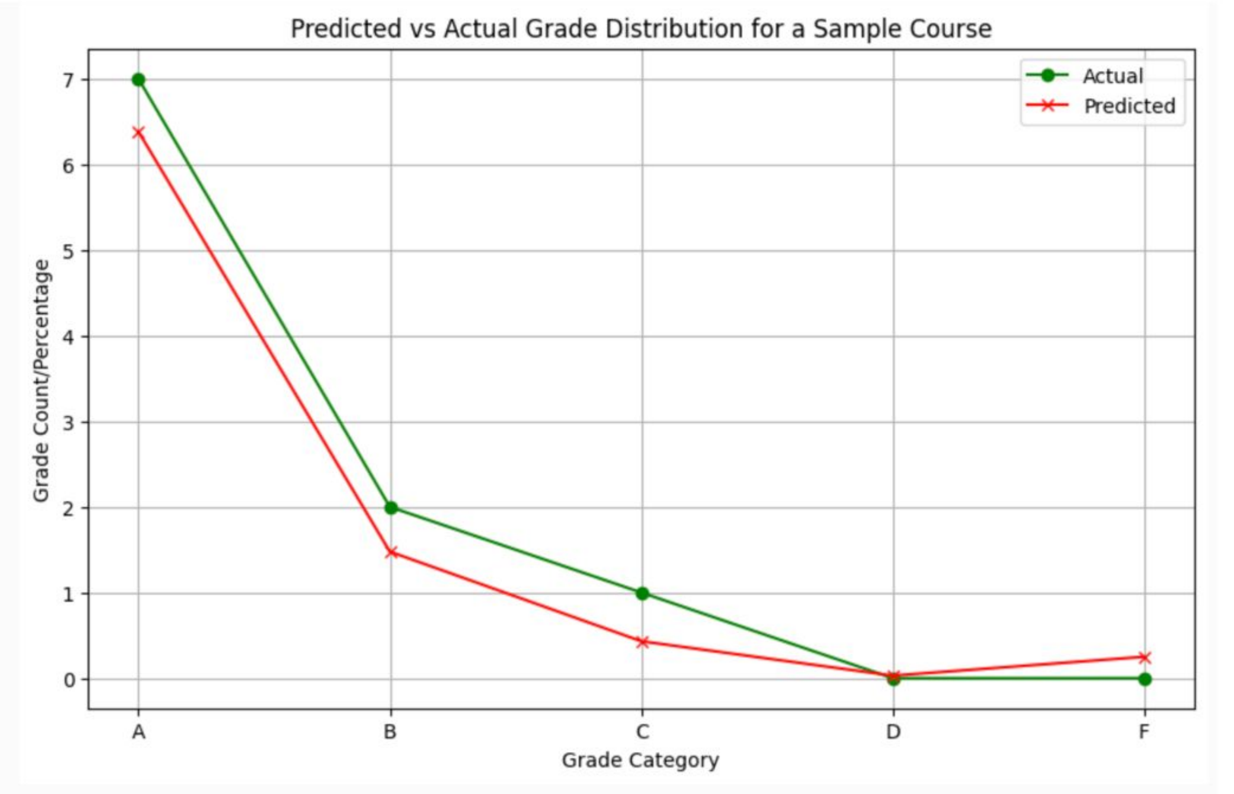
Draw Conclusions from Predictions and Inference

Visualization	Predictions	Actual Outcome	Conclusion
# 1	Higher grades tend to decrease as course levels increase.	Higher grades tend to increase as course levels increase.	Due to the smaller class sizes and the fact that higher-level courses are often taken by MS and PhD students specializing in the area, the average grades in these courses actually increase.
# 2	Student registrations increase with higher instructor ratings, while withdrawal rates decrease as ratings improve.	Withdrawal rates lower as the ratings are higher, and registration rates are higher for professors with higher ratings.	Students see RMP pages to determine if they should register for a course or not. When students register for a course without a professor assigned, they might decide to stay or withdraw after seeing their RMP rating.
#3	The percentage of CS classes conducted online was greater than the number of ME and IE classes post covid.	The percentage of CS classes conducted online was greater than the number of ME and IE classes post covid.	The format of the class and the department of the class are correlated. Therefore they would be a good features to include when training our machine learning model.

Draw Conclusions from Predictions and Inference

	Predictions	Actual Outcome	Conclusion
Model # 1: Grade Distribution Predictor	We can use class data and RMP ratings to predict future grade trends.	Class Size had the highest impact; RMP ratings showed moderate correlation with grades. Semester impact was minimal.	Using previous grade distributions, we established a strong connection to predict future trends. RMP ratings and surprising Class Size influence were key.
Model # 2: Retention Rate Predictor	Use class data and RMP ratings to predict future retention rates.	We've achieved moderate accuracy. Class Level, Instructor Ratings had the most influence on retention. In progress.	We want to experiment with more features like grade distribution and perform feature engineering for better accuracy.

Reports, Decisions, and Solutions



Reports, Decisions, and Solutions (cont.)

- We can inform students on what courses are preferred to take over others.
- Our results provide a solution for students that don't know what professor to choose and/or what courses to take.
- We addressed RMP bias by acknowledging limitations in the analysis, ensured data completeness where possible, and proposed future inclusion of more objective metrics.
- We used a Random Forest Regressor for multi-target prediction of grade categories, optimized hyperparameters, and evaluated performance with MSE.

Thank you!

Questions?