Regression Analysis of College Exams

Josh Grimes

Business Problem

- 62% of graduates go to college
  - 27.1% Psychological Distress
  - 7.3% Burnout
- Our job to prepare them

Goal

- Help prepare high school seniors for the rigors of college with data-supported insights to predicting exam scores

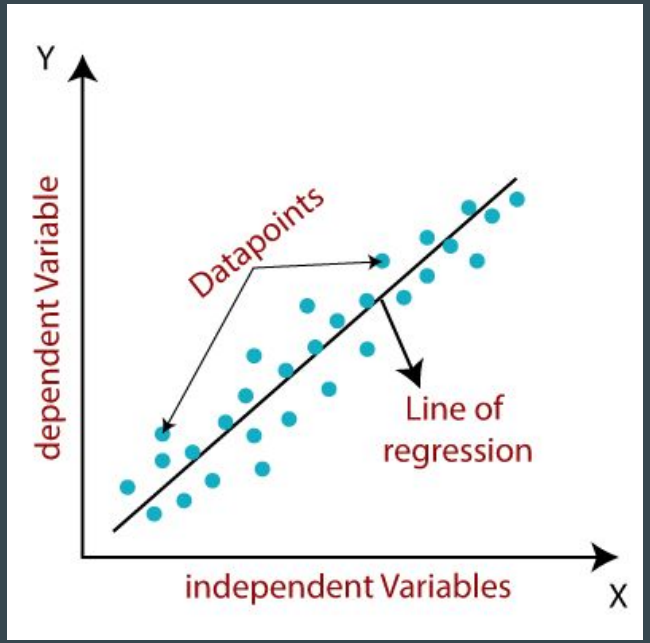
Data

- 6,607 records
- 20 features
- Csv
- Target: Exam Score

Attribute	Description
Hours_Studied	Number of hours spent studying per week.
Attendance	Percentage of classes attended.
Parental_Involvement	Level of parental involvement in the student's education (Low, Medium, High).
Access_to_Resources	Availability of educational resources (Low, Medium, High).
Extracurricular_Activities	Participation in extracurricular activities (Yes, No).
Sleep_Hours	Average number of hours of sleep per night.
Previous_Scores	Scores from previous exams.
Motivation_Level	Student's level of motivation (Low, Medium, High).
Internet_Access	Availability of internet access (Yes, No).
Tutoring_Sessions	Number of tutoring sessions attended per month.
Family_Income	Family income level (Low, Medium, High).
Teacher_Quality	Quality of the teachers (Low, Medium, High).
School_Type	Type of school attended (Public, Private).
Peer_Influence	Influence of peers on academic performance (Positive, Neutral, Negative).
Physical_Activity	Average number of hours of physical activity per week.
Learning_Disabilities	Presence of learning disabilities (Yes, No).
Parental_Education_Level	Highest education level of parents (High School, College, Postgraduate).
Distance_from_Home	Distance from home to school (Near, Moderate, Far).
Gender	Gender of the student (Male, Female).
Exam_Score	Final exam score.

Regression Modeling

- Which feature(s) are important to predict continuous targets

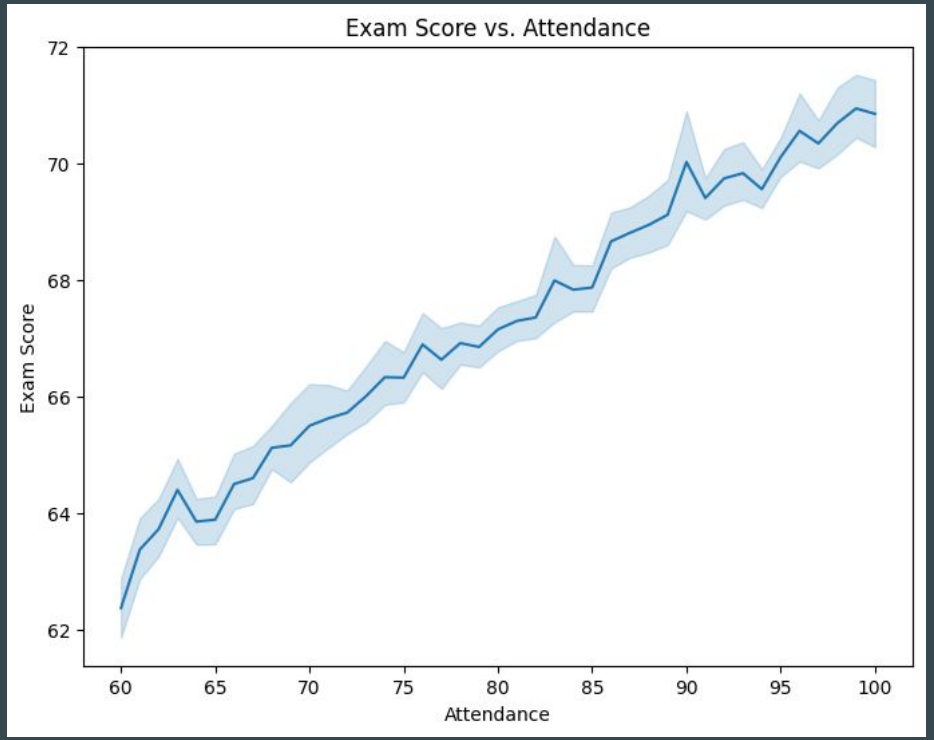


Metrics

- Mean Absolute Error (MAE): average size of mistakes → LOWER = BETTER
- R-Squared = Explained Variance (%)

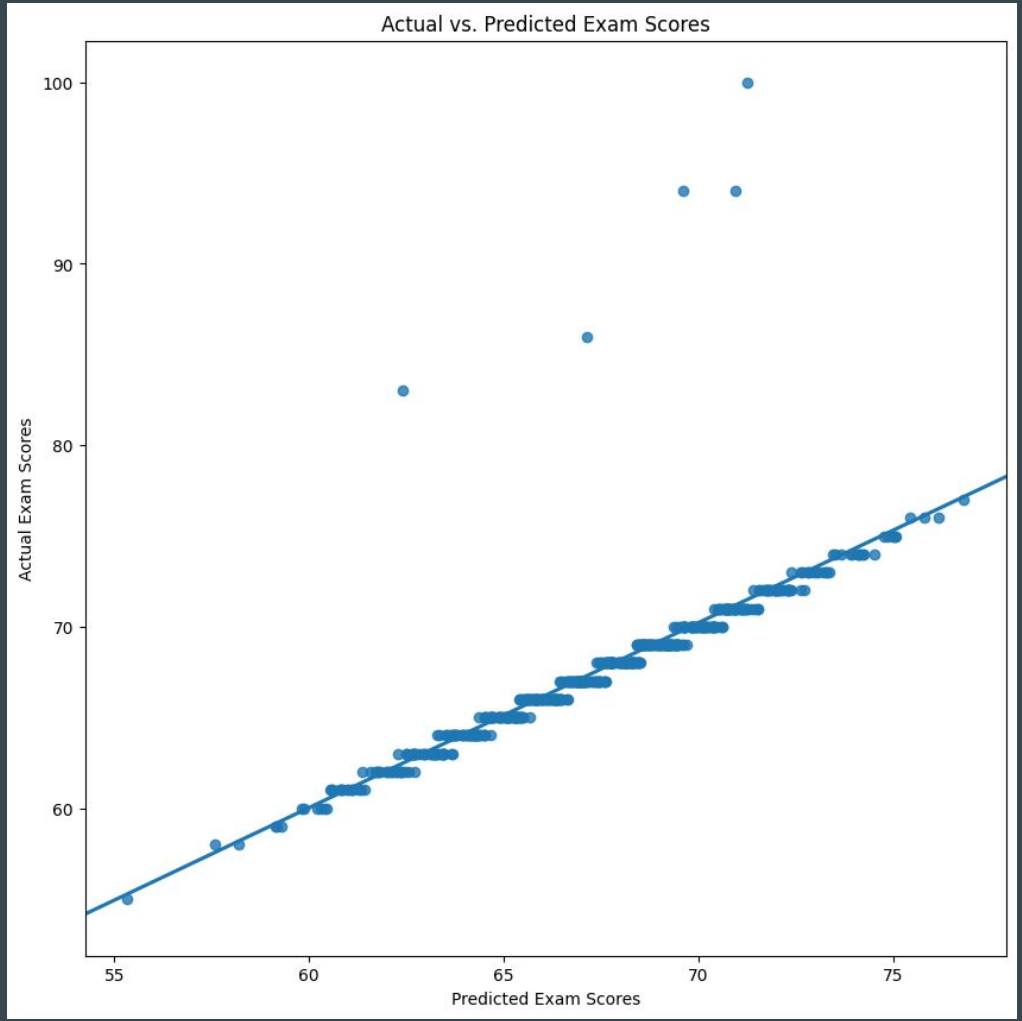
Baseline Model

- Attendance vs. Exam Score
- MAE = 2.073
- R-squared = .331



Final Model

- MAE = .447
- R-Squared = .729



Recommendations

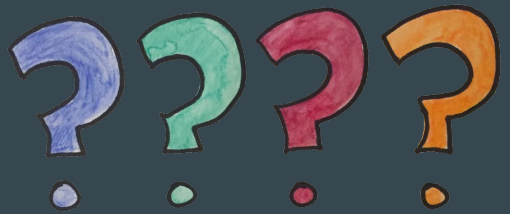
- Go to class
- Study
- Start off strong
- Get help
- Choose good friends
- Stay Home
- Exercise
- Extracurriculars

Feature	Coefficient	Reference Feature
Attendance	2.31	None
Hours Studied	1.79	None
Previous Scores	0.72	None
Tutoring Sessions	0.61	None
Positive Peer Influence	0.49	Negative Peer Influence
Close to Home	0.49	Far from Home
Internet Access	0.25	No Internet Access
Neutral Peer Influence	0.24	Negative Peer Influence
Physical Activity	0.24	None
Extracurriculars	0.24	No Extracurriculars
Moderately Close to Home	0.23	Far from Home
Parent Education (Post-Grad)	0.2	Parent Education (College)
Family Income (Medium)	-0.24	Family Income (High)
Motivation Level (Medium)	-0.24	Motivation Level (High)
Teacher Quality (Medium)	-0.25	Teacher Quality (High)
Parent Education (High School)	-0.25	Parent Education (College)
Teacher Quality (Low)	-0.3	Teacher Quality (High)
Learning Disability	-0.3	No Learning Disability
Motivation Level (Low)	-0.46	Motivation Level (High)
Family Income (Low)	-0.48	Family Income (High)
Access to Resources (Medium)	-0.49	Access to Resources (High)
Parental Involvement (Medium)	-0.5	Parental Involvement (High)
Parental Involvement (Low)	-0.79	Parental Involvement (High)
Access to Resources (Low)	-0.81	Access to Resources (High)

Next Steps

- **New Features:** diet?
- **New Population:** high school?
- **Application?**
- **Explore:** Tutoring Sessions?

Thank You




JOSHUA GRIMES
Data Scientist
919-273-1520
josh.g.grimes@gmail.com
Fayetteville, North Carolina
https://www.linkedin.com/in/joshua-g-grimes/