

EXAM PERFORMANCE and PREDICTION

Data Science Project Using Linear Regression



PRESENTED BY

Khushveer Kaur

Ramcharan

Samatha Sri

Presented To :

Mr. Devansh

About the Project

The role of life style and Habits in students day to day to life :-

Lifestyle and habits are the daily choices and routines that shape how a student lives, studies, and takes care of their health.

Role in a Student's Life:

- * **Improves learning:** Good habits like regular study and time management help students understand lessons better.
- * **Maintains health:** Healthy eating, enough sleep, and exercise keep the body and mind active.
- * **Builds discipline:** Following a routine teaches responsibility and self-control.
- * **Reduces stress:** Balanced habits, including rest and hobbies, help manage pressure.
- * **Shapes personality:** Positive lifestyle choices develop confidence and good character.

The Problem & The Goal :-

The Challenge (Problem Statement) :

- * **Student Success:** Academic performance is crucial, but it's influenced by more than just time spent in class.
- * **The Question:** Which non-academic factors (like sleep, diet, and screen time) have the strongest **linear relationship** with exam scores
- * **Project Goal:** To build a reliable **Linear Regression model** that predicts a student's `exam score` based on their lifestyle and habits.
- * **Key Variables Features ():** Study hours, Attendance, Sleep hours, Social media use, Mental health , etc.

Methodology: Linear Regression:-

- **What is Linear regression?**

- **Definition:** A statistical model used to predict a continuous target variable () based on one or more predictor variables (). It assumes a straight-line relationship between the variables.

The model finds coefficients (weights) for each feature:

- **Coefficients ():** Tell us the size and direction of the impact of each factor on the Exam Score.
- **Training & Testing**
 - **Split:** We divided our 1000 records into:
 - **Training Set (80%):** Used to teach the model habits to scores.
 - **Testing Set (20%):** Used to evaluate the model's performance on unseen data.

Exploratory Data Analysis (EDA) - Initial

Insights

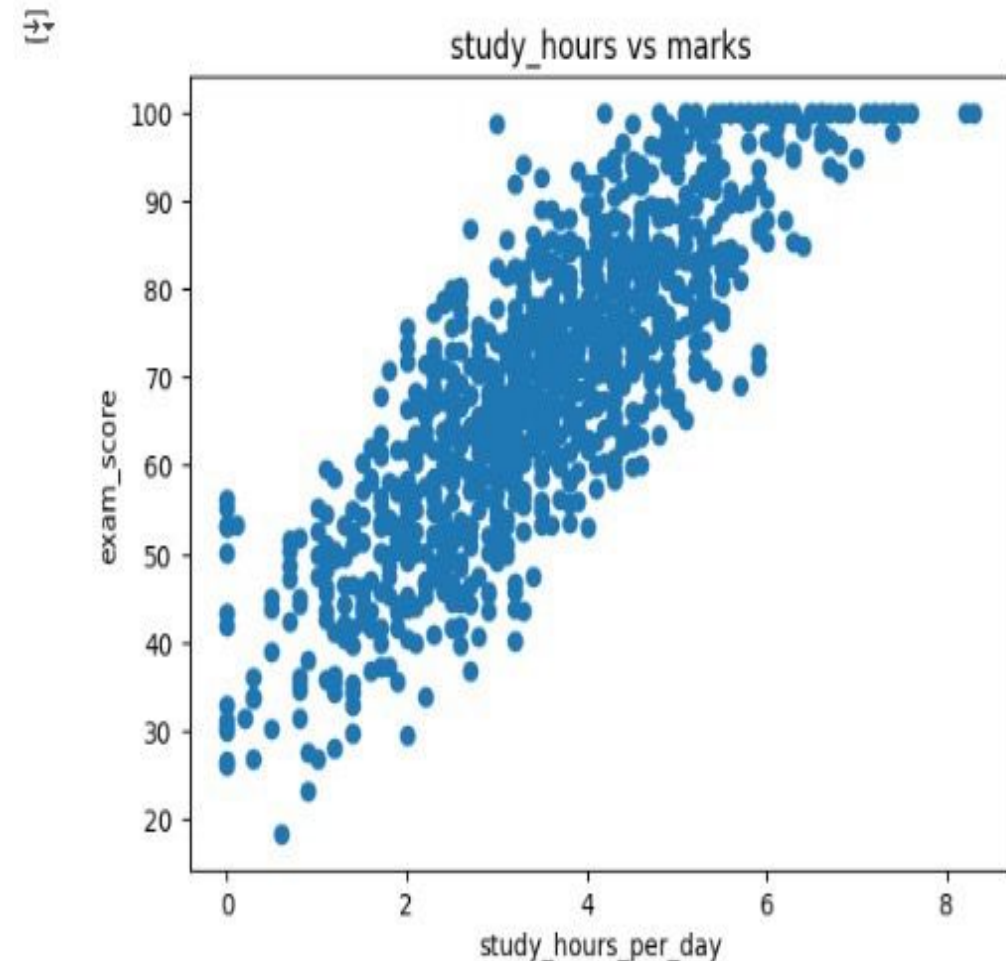
1. Study Hours vs. Exam Score:

- This scatter plot shows the relationship between the number of hours students study per day and their exam scores.

• Key Observations

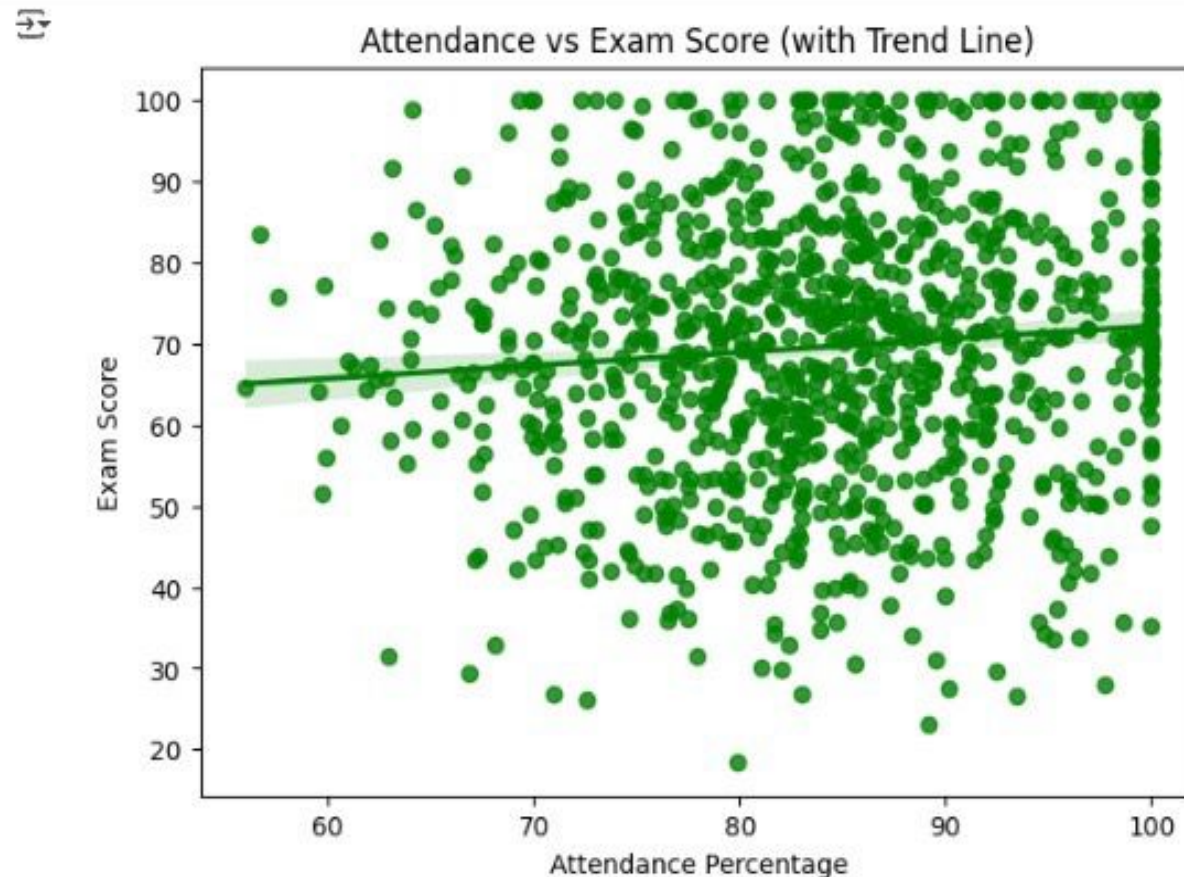
- There is a strong positive correlation between study hours and exam score.
- As study hours increase, exam scores also tend to increase.
- Most students who study 0-2 hours score between 20-60 marks.
- Students who study 3-6 hours commonly score 60-90 marks.
- Beyond 6 hours, many students reach the higher score range (90-100).

```
# visualizing the data
plt.scatter(df['study_hours_per_day'],df['exam_score'])
plt.xlabel('study_hours_per_day')
plt.ylabel('exam_score')
plt.title("study_hours vs marks")
plt.show()
```



Attendance vs Student Exam Scores :-

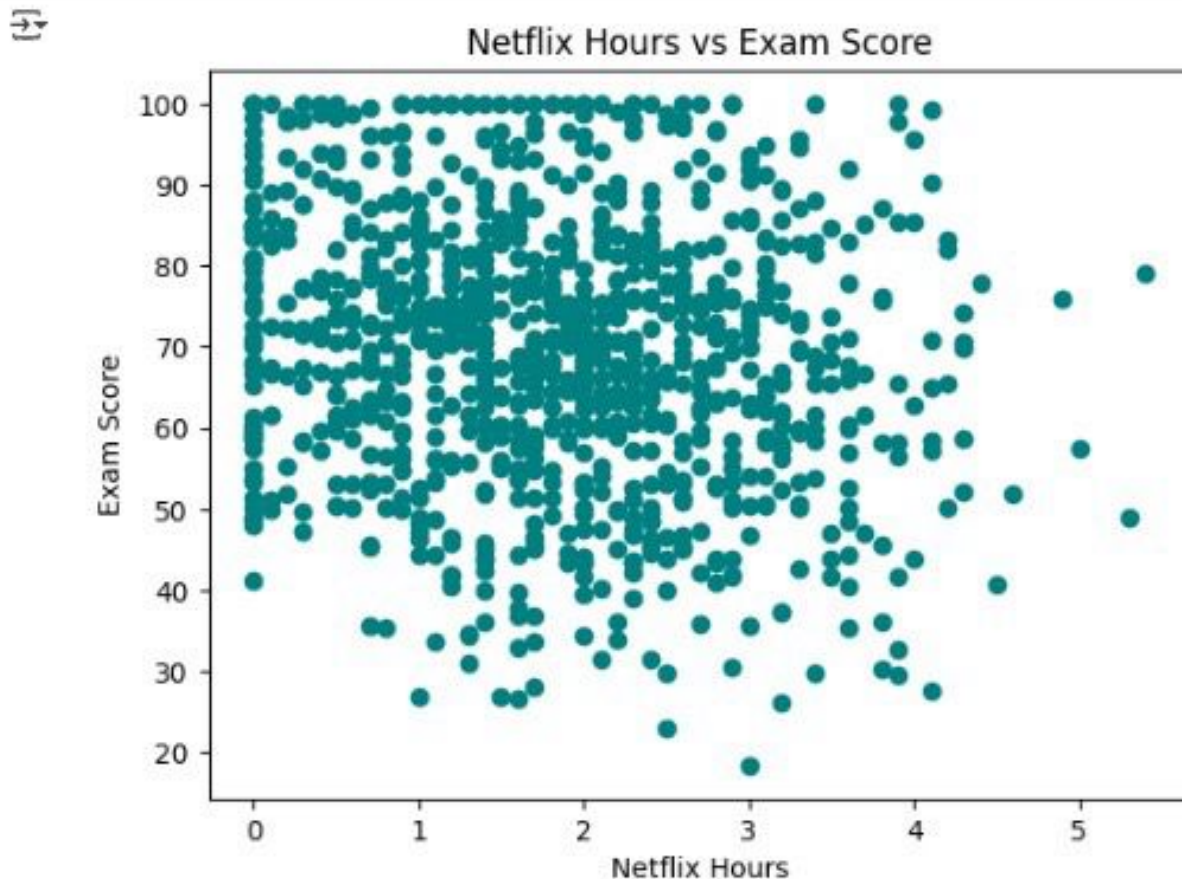
```
plt.figure(figsize=(7,5))
sns.regplot(x='attendance_percentage', y='exam_score', data=df, color='green')
plt.title("Attendance vs Exam Score (with Trend Line)")
plt.xlabel("Attendance Percentage")
plt.ylabel("Exam Score")
plt.show()
```



- It Shows relationship between attendance % and exam scores.
- Trend line is almost flat means very weak correlation.
- High attendance does not guarantee high marks.
- Marks vary a lot at all attendance levels.
- Attendance alone is not a strong predictor of performance.

Exam score vs Netflix Hours

```
plt.scatter(df['netflix_hours'], df['exam_score'], color='teal')  
plt.title("Netflix Hours vs Exam Score")  
plt.xlabel("Netflix Hours")  
plt.ylabel("Exam Score")  
plt.show()
```



- This graph shows how many hours students watch Netflix and their Exam score.
- The dots are spread everywhere with no clear pattern.
- No fixed pattern is seen in the graph of students who watch Netflix and who does not watch Netflix.

Sleep Metric

Duration (Quantity)

Consistency

Bedtime

The "All-Nighter"

Relationship to Exam Score

Strong Positive Correlation.

Students who meet or exceed recommended sleep (typically **7-9 hours** for adults/college students) consistently achieve higher GPAs and better test scores.

Highly Predictive. Students with regular, consistent sleep and wake times (even if total duration is the same) perform significantly better than those with erratic schedules.

A Surprising Cutoff. Studies (like one from MIT) found that going to bed after a certain personal threshold (e.g., 2:00 AM) often resulted in lower grades, *even if the total hours of sleep were the same.*

Negative/No Correlation. The amount of sleep on the *single night before the exam* is not the most important factor. If you consistently sleep poorly, one good night won't help. Pulling an all-nighter is

Key Mechanism

Sleep deprivation impairs **attention, focus, and long-term memory** retrieval, leading to lower performance.

An irregular schedule disrupts the **circadian rhythm**, leading to fragmented sleep quality and chronic daytime sleepiness.

This suggests that the timing of sleep, which controls deep (REM) sleep cycles, is critical for cognitive restoration.

After about 20 hours of wakefulness, your performance is akin to being legally impaired. You won't use the memorized information effectively.

Sleep hours vs Average Exam scores



This graph shows the relationship between the number of hours students sleep each day and their average exam scores.

Key Observations

As sleep hours increase from 3 to around 7 hours, the average exam score shows a gradual improvement.

Students who sleep between 6 to 8 hours generally perform the most consistently, with scores mostly between 65-75.

At very high sleep levels (9-10 hours), scores show more variation:

Some students score very high (Above 85).

: Results and Key Findings:-

Model Performance:

- * Linear Regression model achieved an **Score of 0.897** on the test data.

- * **Interpretation:** This is an **excellent result!** It means our features (study habits, lifestyle, etc.) explain almost **90%** of the variability observed in the student exam scores. The model is highly predictive.

Live Prediction Example:

- * **Input Data:** A hypothetical new student with moderate habits (e.g., 4 study hours, 7 sleep hours, high mental health).

- * **Predicted Exam Score: 65.22**

Actionable Insights:

Academic Commitment is Critical:

Study hours and attendance are the primary drivers of high scores.

Health Matters:

Factors like sleep hours, diet quality, and mental health rating hold significant positive weight in the prediction, indicating that holistic well-being directly contributes to academic success.

Conclusion and Future

Work

Conclusion :

Our data science project successfully created an accurate and **interpretable** model (Linear Regression) that quantifies the relationship between student lifestyle data and their performance.

Future Work

- Non-Linear Models:** Explore complex models (like **Random Forest** or **Gradient Boosting**) to see if we can push the score even higher by capturing non-linear relationships.
- Feature Importance Analysis:** Use techniques to formally determine the single most and least impactful features to provide more targeted recommendations.
- Categorical Deep Dive:** Analyze the specific coefficients for categorical features (e.g., is 'Master' parental education much more impactful than 'High School')

THANK YOU