SIMPLE LINEAR REGRESSION

MANASI ODASSERY

Delivery Time Prediction Model

- Define the Problem: Predict delivery time based on sorting time and predict salary based on years of experience.
- Collect and Prepare Data: Loaded and displayed the data. Next we we checked to handle any missing values or outliers if necessary.
- Explore the Data: Performed exploratory data analysis (EDA) using various plots to understand the distributions and relationships within the data. Scatter plots can be particularly useful to visualize the relationship and to identify any patterns or anomalies in the data.
- Model Building: Trained simple linear regression models.

Purpose

- In logistics, timing is critical. Optimizing delivery times can drastically improve operational efficiency and customer satisfaction.
- Develop a predictive model to forecast delivery times based on the sorting time required for packages, allowing for better scheduling and resource allocation.

Data Collection and Description

Analysis based on a dataset of 21 data points

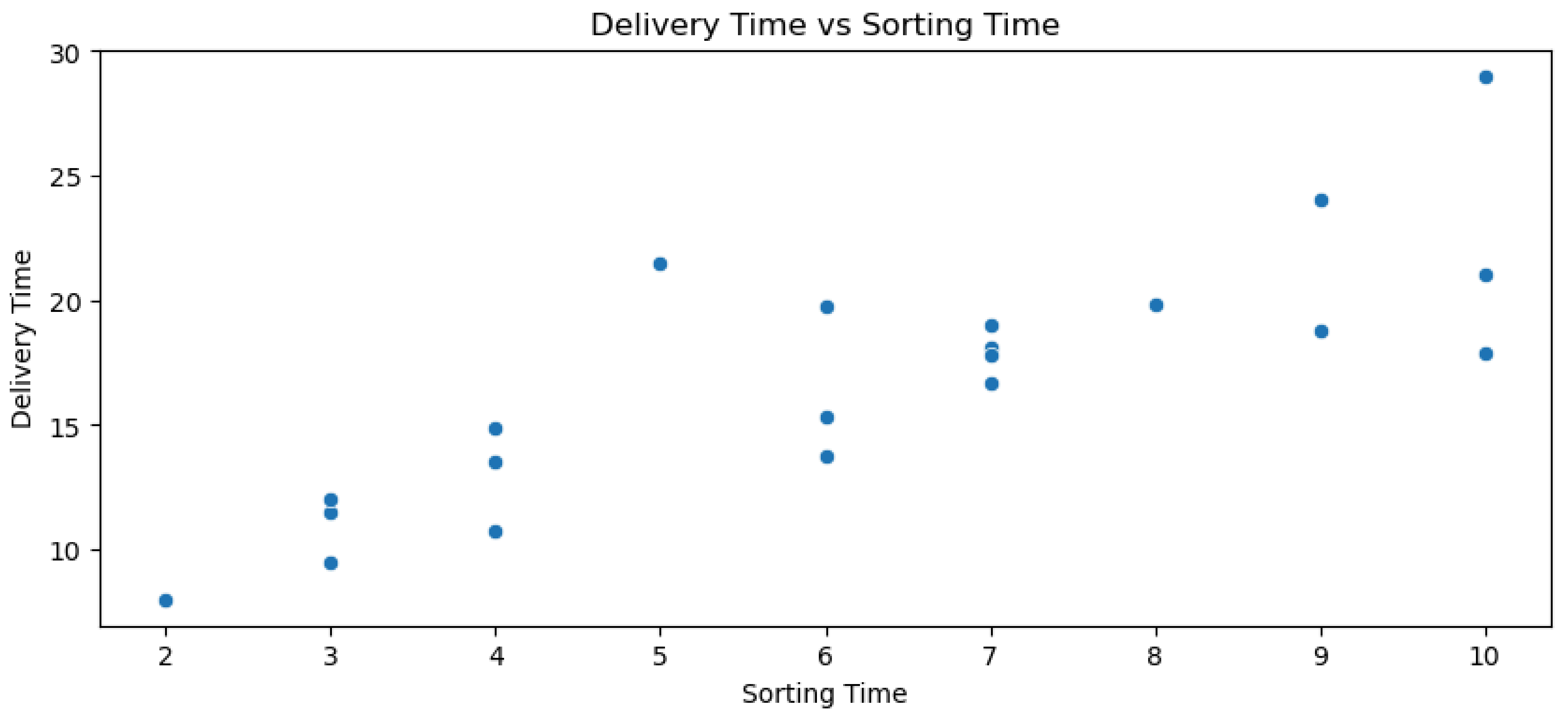
Data Source

Compiled from internal logistics records covering the past year.

Description

- Sorting Time: Time taken to sort packages at the facility (in minutes).
- Delivery Time: Actual time taken from dispatch to delivery (in minutes).

Exploratory Data Analysis (EDA)



Scatter Plot Analysis

- There's a clear positive correlation indicating that as sorting time increases, so does delivery time. This trend suggests that more complex or larger volume sorting tasks may extend the delivery periods.

- The data shows variability, especially with some points at higher sorting times showing significantly increased delivery times, which might indicate occasional operational delays or inefficiencies.

- While a linear relationship is evident, the spread of points suggests that a simple linear model might not capture all dynamics of the data, particularly at the extremes.

Model Evaluation

It highlights the model's effectiveness and potential areas for improvement.

Evaluation

After training, the model was tested on the remaining 20% of the data to evaluate its predictive power and to prevent overfitting.

Mean Squared Error (MSE):

- Value: 14.0467
- Interpretation: This metric indicates the average squared difference between the observed actual outcomes and the values predicted by the model. An MSE of 14.0467 suggests that, on average, the model's predictions deviate from the actual observed values by a square root of approximately 3.75 minutes.

R-Squared (R^2):

- Value: -1.0208
- Interpretation: The R^2 value is a statistical measure that represents the proportion of the variance for a dependent variable that's explained by an independent variable or variables in a regression model. Typically, an R^2 value ranges from 0 to 1, where closer to 1 indicates better model performance. However, an R^2 value of -1.0208 is highly unusual and suggests that the model performs worse than a simple horizontal line representing the mean of the dependent variable (delivery time)

Business Impact and Recommendations

Business Impact

- Operational Efficiency and Cost Savings: Improved prediction of delivery times can enhance logistical efficiency, allowing for better resource allocation and scheduling, which may reduce operational costs.

- Enhanced Customer Satisfaction: More accurate delivery forecasts can significantly improve customer satisfaction by providing reliable service, setting the company apart from competitors.

Challenges and Recommendations

- Model Limitations: Given the negative R^2 value, the current model may not adequately capture the dynamics of delivery times, suggesting the need for more complex modeling approaches or additional explanatory variables.

- Strategic Enhancements: To improve model accuracy, incorporate advanced analytical techniques and enrich the dataset with more relevant factors like traffic and weather conditions.

Salary Based on Experience

- Define the Problem: Build a model to predict salary increases based on years of experience, which will help in strategic salary planning and maintaining market competitiveness.

- Collect and Prepare Data: Loaded and displayed the data. Next we we checked to handle any missing values or outliers if necessary.

- Explore the Data: Performed exploratory data analysis (EDA) using various plots to understand the distributions and relationships within the data. Scatter plots can be particularly useful to visualize the relationship between years of experience and salary, and to identify any patterns or anomalies in the data.

- Model Building: Trained simple linear regression models.

Purpose

- In human resource management, compensation alignment is crucial. Optimizing salary structures based on years of experience can significantly improve financial planning and employee satisfaction.
- Develop a predictive model to forecast salary hikes based on the years of experience employees hold, enabling better budgeting and equitable salary distribution. This approach aids in maintaining a competitive edge in talent acquisition and retention by ensuring fair and transparent compensation practices.

Data Collection and Description

Based on a dataset of 30 data points, providing a robust sample for assessing the impact of years of experience on salary.

Data Source

Compiled from internal HR records covering the past decade.

Description

- Years of Experience: Time in years that employees have worked in the industry.
- Salary: Actual annual salary of the employees in USD.

Exploratory Data Analysis (EDA)

Scatter Plot Analysis

- There's a distinct positive correlation showing that as years of experience increase, so does the salary. This trend suggests that with greater experience, employees tend to earn higher wages, reflecting their accumulated knowledge and skills.

- The scatter plot shows that with every additional year of experience, there's a general increase in salary, though the rate of increase varies. This variation might indicate differences in career progression paths, roles, or industries.

- A linear relationship is evident between the years of experience and salary, making linear regression a suitable method for modeling these data. However, the variability suggests that for higher accuracy, considering additional factors or a non-linear model might better capture salary dynamics.

Model Evaluation

It highlights the model's effectiveness and potential areas for improvement.

Evaluation

After training, the model was tested on the remaining 20% of the data to evaluate its predictive power and to prevent overfitting.

Mean Squared Error (MSE):

- Value: 49838096.85598939
- Interpretation: This metric quantifies the average squared difference between the actual salaries and the predicted values by the model. The MSE of 49838096.86 suggests that, on average, the model's predictions deviate from the actual observed salaries by a square root of approximately 7059.52, indicating the typical error in salary prediction.

R-Squared (R^2):

- Value: 0.9024461774180497
- Interpretation: The R^2 value measures how well the observed outcomes are replicated by the model, based on the proportion of total variation of outcomes explained by the model. An R^2 of 0.9024 implies that the model explains approximately 90.24% of the variance in salaries from years of experience, showcasing a high level of predictive power.

Business Impact and Recommendations

Business Impact

- Strategic HR Management and Budget Planning: Accurate predictions of salary based on years of experience can significantly enhance budget planning and resource allocation within HR, ensuring that salary increments are well-aligned with industry standards and employee expectations.
- Talent Acquisition and Retention: By ensuring competitive salary offerings that reflect market trends and employee experience, the organization can improve its ability to attract and retain top talent, thereby reducing turnover and associated training costs.

Challenges and Recommendations

- Model Limitations: Although the model demonstrates high predictive accuracy ($R^2 = 0.9024$), there is always room for improvement. The model currently only includes years of experience as a predictor, which may overlook other factors influencing salary variations such as role, industry, and geographic location.
- Strategic Enhancements: To further improve the accuracy and applicability of the model, consider incorporating additional variables including more detailed employee data such as education level, certifications, job role, performance ratings, etc. and advanced analytical techniques to explore more sophisticated modeling techniques like multiple regression or machine learning to handle complex interactions between multiple factors affecting salaries.

THANK YOU