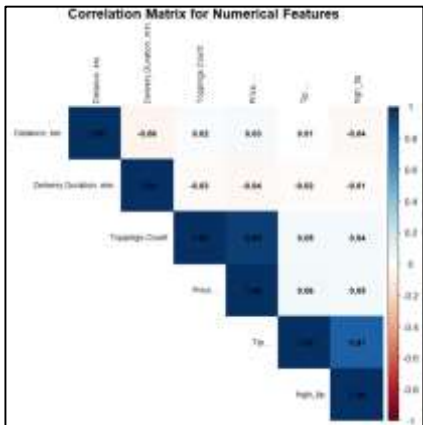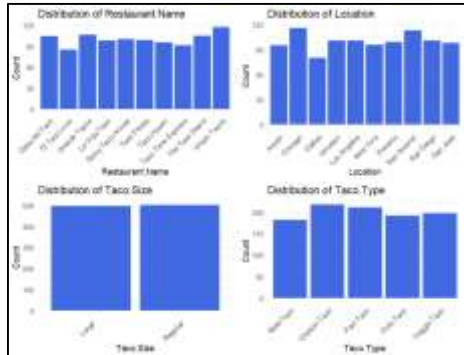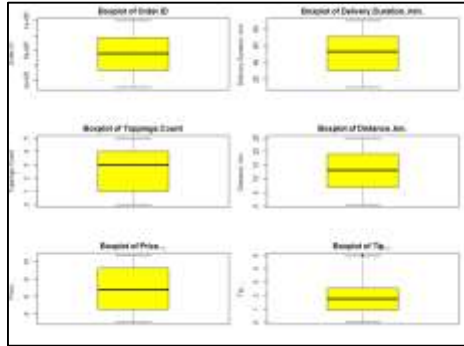# BUSINESS INTELLIGENCE

Predicting Tips in Taco Deliveries using Regression Models

Gülce Çelik

Supervisor: Prof. Giovanna Maria Dimitri

Academic Year: **2024–2025**

PROJECT OVERVIEW & OBJECTIVE

Objective:

- Predict **Tip Amount ($)** using features like taco type, delivery time, price, location, etc. Evaluate **3 models**: Linear Regression, Decision Tree, Neural Network

Dataset Summary:

- 1,000 taco orders (2024–2025)

- Features: price, toppings, delivery duration, weekend flag, restaurant info, etc.

- Target: **Tip ($)**

Key Preprocessing Steps:

- Handled outliers using **Z-score and IQR**

- One-hot encoding for categorical features

- Z-score normalization for neural network

- Feature selection tested in **two different sets**

Key Preprocessing Steps

•Visualized distributions using **boxplots**, **Q-Q plots**, and **bar charts**

•Detected and **capped outliers** via **Z-score** and **IQR** methods

•Checked **missing values** – none found

•Built **correlation matrices** (numeric & mixed-type) to assess relationships

•Converted **categorical** and **boolean** variables using **one-hot encoding**

•Dropped time-related columns (Order/Delivery Time)

•Applied **Z-score normalization** for Neural Network

•Designed and tested **two feature sets** for model comparison
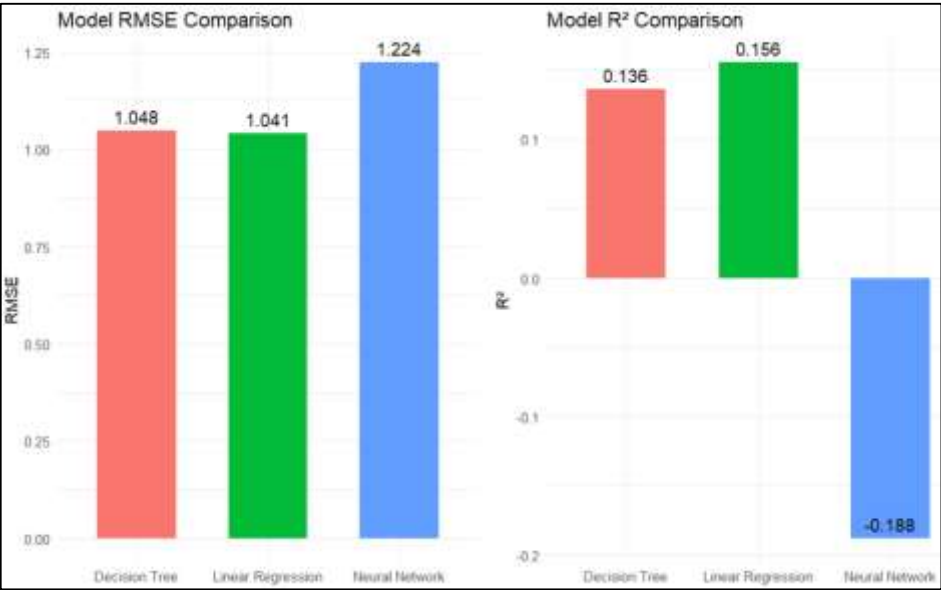
MODEL RESULTS & COMPARISON

Regression Model Performance (5-Fold CV)

Model	Avg. RMSE	Avg. R^2
Linear Reg.	~1.04–1.05	~0.14–0.16
Decision Tree	~1.05–1.07	~0.11–0.13
Neural Network	~1.22–1.25	~−0.19 to −0.23

Best Model: Linear Regression (stable & interpretable)

Neural Network: Worst performance (low correlation, potential overfitting)

Insight: Tip shows **weak correlation** with most features

INSIGHTS,LIMITATIONS&FUTURE WORK

🔍KEY INSIGHTS
-Tip behavior is weakly correlated with available data
-Linear regression performed best
-Possibly influenced by external factors:
   -Customer mood
   -Driver behaviour
   -Package condition
   -Delivery experience

📊ADDITIONAL TEST
-Predicting Price instead of Tip
   -Models achieved $R^2 > 0.90$
   -Confirms the modeling pipeline is valid
-Tip prediction failed due to limited feature signals

💡FUTURE RECOMMENDATIONS
-Include other possible external factors to the data set
-Collect behavioral and psychological data
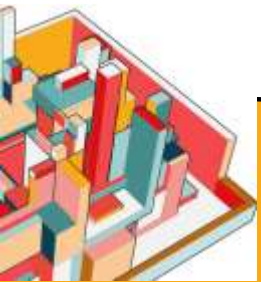
THANKS

PREDICTING TIPS FROM TACO DELIVERY DATA

Gülce Çelik

Matriculation:157341

Supervisor: **Prof. Giovanna**

Maria Dimitri