


# AN ANALYSIS AND PREDICTION OF EMO'S FIRST ELECTRIC MOPED OWNERSHIP



# Today's Agenda:

- Client's Problem
- Process
- Insights
- Results

# Client's Problem

---

**EMO** is a manufacturer of motorcycles. The company successfully launched its **first electric moped** in India in **2019**. The product team knows how valuable **owner reviews** are in making improvements to their mopeds.

**Unfortunately**, they often get reviews from people who **never owned** the moped. They don't want to consider this feedbacks, so would like to find a way to **identify** reviews from these people. They have obtained **data** from other, where they know if the reviewer owned the moped or not.

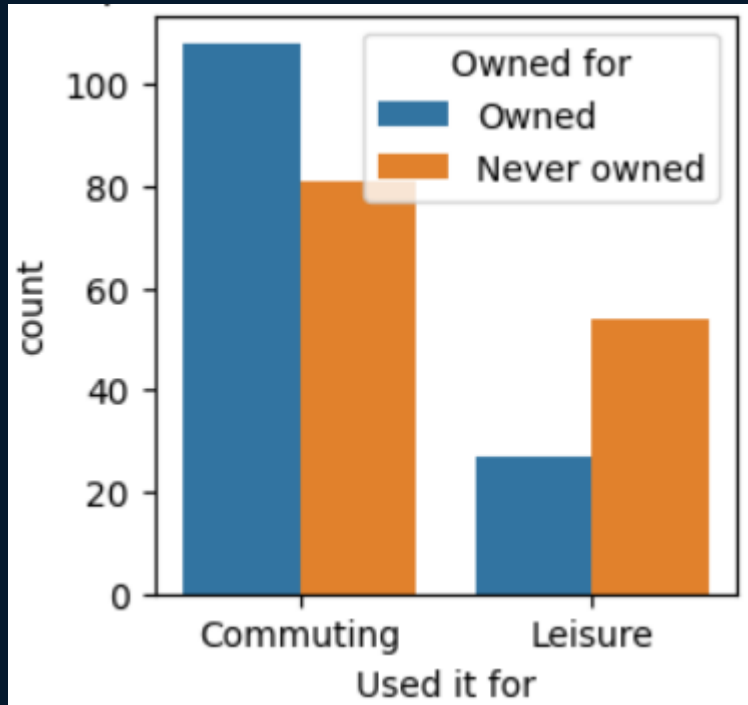
The **manager** asks me to **predict** which reviews come from people who have never owned the moped before and for those who does.



# Process

- 1 Data Validation
- 2 Exploratory Analysis
- 3 Model Fitting
- 4 Model Evaluation

# Insights



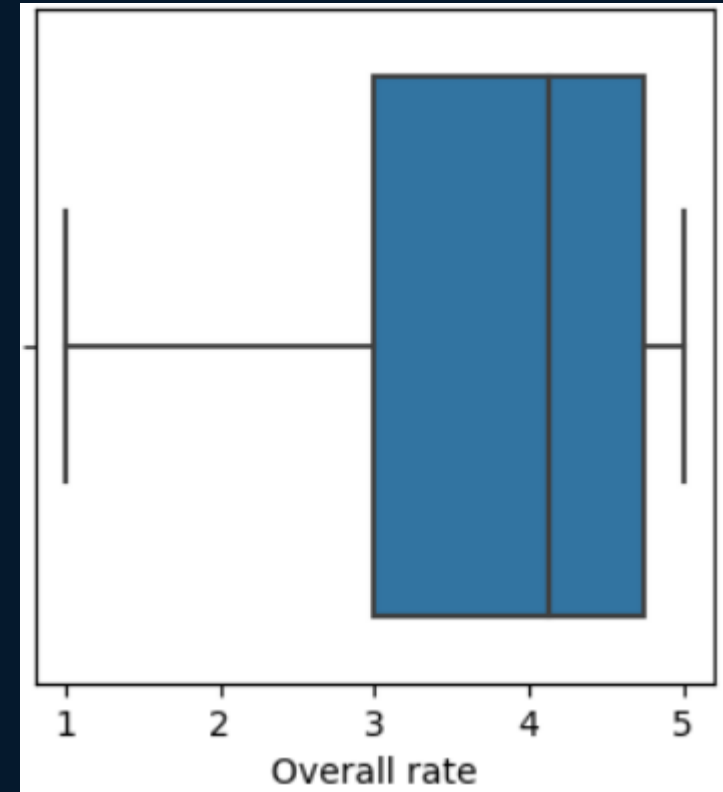
A **count plot** for users who uses the moped for **commuting** or for **leisure** activities.

It can be seen from the plot that most users who **owned the moped uses it mainly for commuting** while the users **who do not own it uses it for doing leisure** activities.

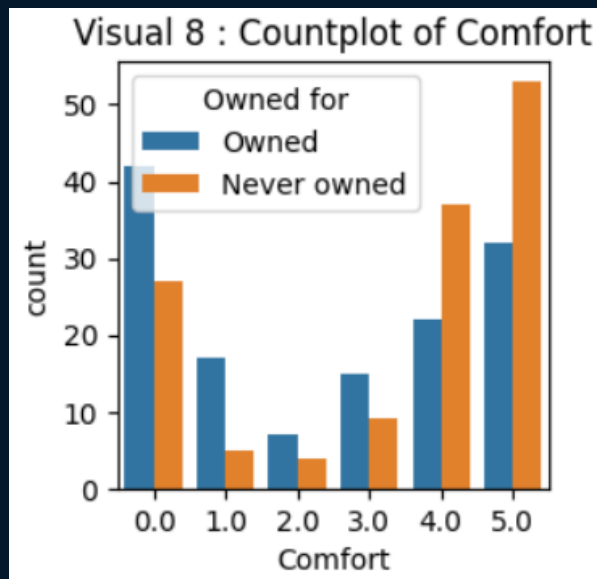
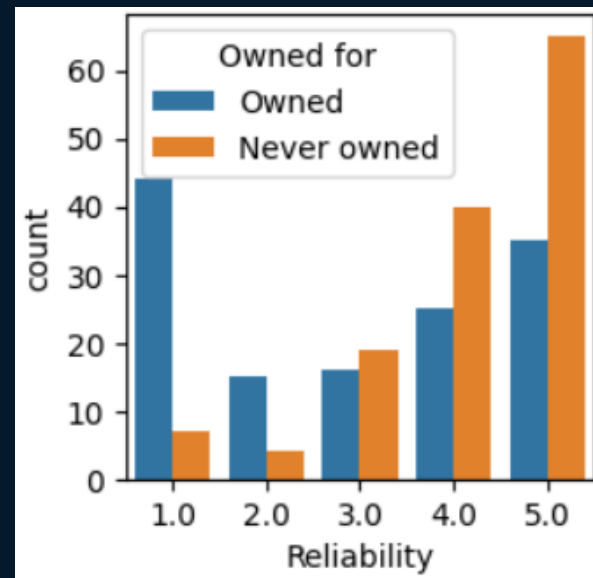
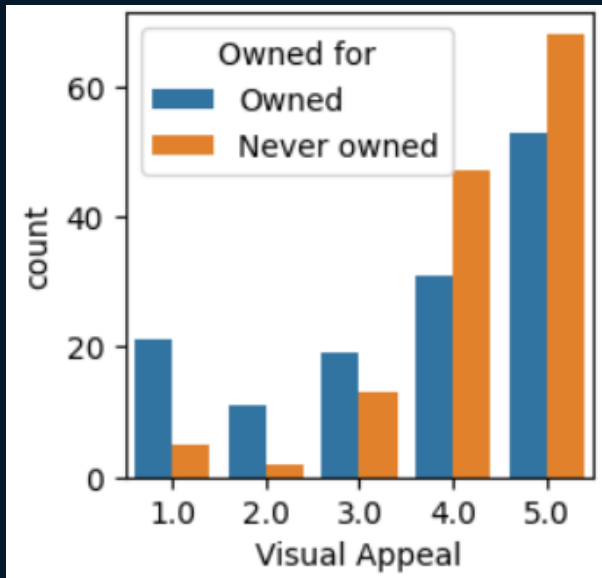
# Insights

A **box plot** of the average rating of the moped.

The plot shows that majority of the ratings are **3 and above** with a **median of higher than 4**. This indicates a good score for their product.



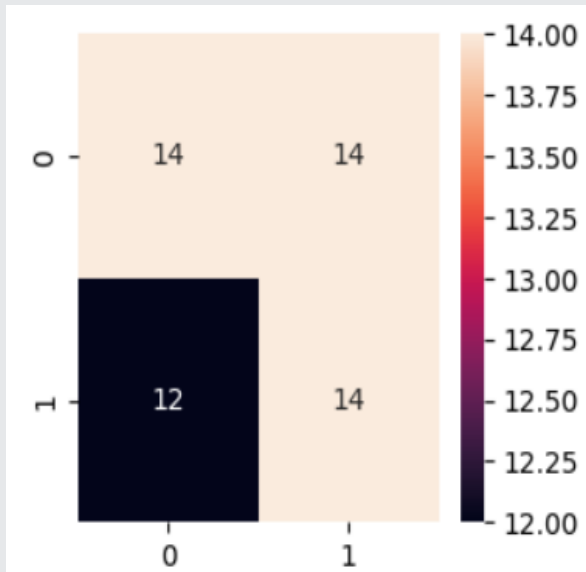
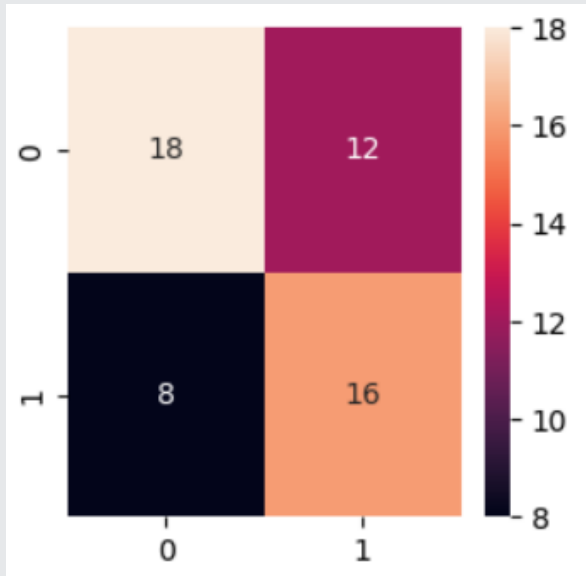
# Insights



These are **count plots** based from the ratings that gave distinction on certain users.

All of these plots shows that most users **who owned a moped gave a low rating** in terms of visual appeal, reliability, and comfort. While this is the **opposite for the ones who never owned** the moped

# Results



The above confusion matrix shows the results of the predictions of **logistic regression** while the one below is the results of the predictions of **decision tree**. Logistic regression predicted 18 True Positive, 16 True Negative, 12 False Positive, and 8 False Negative. Decision tree predicted 17 True Positive, 13 True Negative, 13 False Positive, and 9 False Negative. **Logistic regression** achieved an **accuracy of 62%** while **Decision tree** achieved an **accuracy of 56%**. Logistic regression has a better way of predicting whether an entry is by an owner or not.





# Thank you!

You can reach out to me at my  
Email: [raimondextervinluan@gmail.com](mailto:raimondextervinluan@gmail.com)  
LinkedIn profile: [link](#)

Check out the workspace [here](#)