

# Chap4\_LuRuiqi.docx

*by* LU RUI QI

---

**Submission date:** 24-Jul-2024 12:48PM (UTC-0700)

**Submission ID:** 2421918821

**File name:** Chap4\_LuRuiqi.docx (15.1K)

**Word count:** 478

**Character count:** 2851

## **Chapter 4: EDA/Initial Results**

### **Case 1: Primary Data Exploratory Data Analysis (EDA)**

**1. Visualizations Traffic Volume Trends:** Line graphs of daily traffic volume trends. **Weather Impact:** Scatter plots to see the relationship between weather parameters (e.g. rain, temperature) and traffic volume. **Accident Statistics:** Bar plots depicting number of accidents on each month or year.

**2. Descriptive Statistics** Summary statistics of traffic volume, weather parameters, accident amounts. Correlation matrix showing what selected parameters are correlated with traffic volume.

**3. Initial Insights** Which is the peak hour of the day for traffic volume and the conditions of traffic during that period. Which month had maximum number of accidents. Insight into weather conditions for that specific month.

**4. Feature Engineering** New columns like **Rainy\_Days**: whether these days have above trace amount of precipitation and **Peak\_Hours** Encoding the weather column and other categorical/descriptive columns.

### **Case 2: Secondary Data Exploratory Data Analysis (EDA)**

**1. Visualizations Ridership trends:** Line plots of daily/monthly/yearly public transport ridership trends. **Service Comparison:** Box plots comparing ridership statistics of different transport types.

**2. Descriptive Statistics** Calculating summary statistics of ridership data, average daily ridership, ridership std dev etc. Inspecting the ridership usage trend over different transport systems.

**3. Initial Insights** What are the top hours for usage of public transport in a day. What are the demand peaks for each transport separately. Correlating ridership data with weather conditions to assess effect of unfavorable weather.

#### 4. Feature Engineering

Creating dictionary for transport descriptions and encoding them. Feature Engineering Finding out two new features "Weekend" and "Holiday" to differentiate the weekdays from weekends and holidays from usual working days It is assumed that holidays would have less vehicular traffic as compared to weekdays.

Creating Lag features as well as Bollinger Band features for capturing the effect of past traffic on present and future traffic

#### Machine Learning (Initial Results)

1. Model Selection Build regression models: first Linear Regression and then Decision Trees predicting the traffic based on Weather and Ridership data using metrics such as Mean Absolute Error (MAE) and R-squared to evaluate the models

2. Initial Observations Weather data and Public transport ridership data have turned out to be highly predictive for the traffic volume Decision Trees have performed better than other models. Reason could be that decision trees are good at capturing non-linearity and the numerous factors like temp, rain, snow etc can have interesting non-linear relations with traffic.

3. Next Steps Tune the model parameters for better performance Include additional features like special events and Holidays to get more information on public transport ridership. Once we have been through the different analyses and visualizations as mentioned above, we establish the insights effectively.

There is a lot that could be done to perform good visualization. Tools to perform the visualization may include Tableau, R (ggplot2) etc. The next step could be to construct powerful predictive models for the ultimate traffic parameter prediction in Malaysia.

ORIGINALITY REPORT

2%

SIMILARITY INDEX

2%

INTERNET SOURCES

0%

PUBLICATIONS

0%

STUDENT PAPERS

PRIMARY SOURCES

1

rgu-repository.worktribe.com

Internet Source

2%

Exclude quotes On

Exclude matches Off

Exclude bibliography On