DATA INTAKE REPORT

1. Introduction:

The G2M Insight for Cab Investment Firm project aims to provide valuable insights and recommendations for a cab investment firm by analyzing data related to cab services, customer demographics, and financial performance of two cab companies: Yellow Cab and Pink Cab. The project involves data exploration, hypothesis testing, and model building to support strategic decision-making.

2. Data Sources:

The project involves four individual datasets covering the time period from 31/01/2016 to 31/12/2018:

1. **Cab_Data.csv:** Details of transactions for two cab companies.
2. **Customer_ID.csv:** A mapping table linking customer demographic details.
3. **Transaction_ID.csv:** A mapping table linking transactions to customers and payment modes.
4. **City.csv:** Contains a list of US cities, their population, and the number of cab users.

3. Third-Party Data:

The analysis is not limited to the provided datasets; third-party datasets relevant to the overall theme and geographical properties of the data may be used. Additional research on the overall cab industry in the US will be conducted and correlated with the trends observed in our dataset.

4. Exploratory Data Analysis (EDA) and Hypothesis Testing

Objective: Explore the data, identify patterns, conduct statistical analysis, and test hypotheses to gain insights into the cab services, customer behavior, and financial performance.

Commencement Date: 12th March 2024

Expected Completion Date: 14th March 2024

5. Interpretation and Recommendation Report

Objective: Interpret findings from the EDA and hypothesis testing phase, derive actionable insights, and provide recommendations for the executive team of the cab investment firm.

Expected Completion Date: 21st March 2024

DATA INTAKE REPORT

6. Model Building and Deployment

Objective: Develop predictive models to forecast demand, optimize pricing strategies, and enhance operational efficiency. Deploy the models for real-time decision support.

Commencement Date: 29th March 2024

7. Technical Implementation

1. **Tools and Libraries**
   - Python programming language
   - Libraries: pandas, numpy, matplotlib, seaborn, statsmodels, scipy
2. **Data Sources**
   - Cab transaction data
   - Customer demographics data
   - City data
   - Holidays data
3. **Data Preprocessing**
   - Merge relevant datasets
   - Handle missing values
   - Convert data types
   - Create new features such as 'Is Holiday' and 'Profit on Trip'
4. **Exploratory Data Analysis**
   - Distribution analysis of numerical columns
   - Company performance analysis by key performance indicators (KPIs)
   - Bivariate analysis to explore relationships between variables
5. **Hypothesis Testing**
   - Test assumptions for parametric and non-parametric tests
   - Perform Mann-Whitney U test for significant differences between Yellow Cab and Pink Cab
6. **Interpretation and Recommendations**
   - Interpret findings from EDA and hypothesis testing
   - Derive actionable insights and recommendations for the executive team
7. **Model Building and Deployment**
   - Develop predictive models for demand forecasting and pricing optimization
   - Deploy models for real-time decision support

DATA INTAKE REPORT

8. Timeline

- Data Intake Phase Completion: 12th March 2024
- EDA and Hypotheses Testing Commencement: 12th March 2024
- EDA and Hypotheses Testing Completion: 14th March 2024
- Interpretation and Recommendation Report Presentation: 21st March 2024
- Model Building and Deployment Commencement: 29th March 2024

By following this timeline, the project aims to deliver actionable insights and deployable models to support the cab investment firm's strategic decisions effectively.

9. Challenges and Considerations:

Identify potential challenges, limitations, and considerations that may impact the analysis. This could include data quality issues, missing values, outliers, or any external factors affecting the cab industry during the specified time period.

This Data Intake Report sets the foundation for a thorough and insightful analysis, providing XYZ with actionable insights to make informed investment decisions in the Cab Industry.