SMU
SINGAPORE MANAGEMENT UNIVERSITY

School of
Computing and
Information Systems

IS483 Project Proposal

Project CDG
Team CDG

Main Track: Business Analytics (BA)

Team Members

Name	ID	Email	Role
Choo Hui Xing	01378884	hxchoo.2018@sis.smu.edu.sg	Back-end Developer
Iris Har Jing Ru	01375564	iris.har.2018@sis.smu.edu.sg	Data Engineer
Ivy Hau Jia Yi	01350274	ivy.hau.2018@sis.smu.edu.sg	UI/UX Designer
Shazarifah Binte Shawal	01373871	shazarifah.2018@sis.smu.edu.sg	Data Analyst
Shermin Tan	01374546	shermin.tan.2018@sis.smu.edu.sg	Project Manager
Sonia Johari	01376980	soniajohari.2018@sis.smu.edu.sg	Front-end Developer

Sponsor and/or Clients

Name	Email	Organization	Department	Role
Gary How	garyhow@comfortdelgro.com	ComfortDelGro Corporation	Group Digital Office	Data Team Lead

1.0 Executive Summary

This project aims to increase efficiency and productivity of ComfortDelgro (CDG) in various aspects such as by gaining a greater understanding of Singapore's current transportation market demands, they can easily identify the different market gaps that they could forward into through visualizing the analysis data on a web-based dashboard. Additionally, with such a tool on hand, CDG can easily segment their customer base to adopt different market supply strategies for their target market.

2.0 **Project Overview**: Business Aspects

CDG is one of the largest land transport companies in Singapore and in the world, with a global shareholder base and outlook. Some of CDG's business units include buses, taxis, trains, and car rental and leasing. The project seeks to provide further market insights using data on consumer behaviour and perception in Singapore transport landscape to help CDG optimise their market supply strategies and fill the gaps in the market.

2.1 Project Description

Our project will cover aspects of building a dashboard that performs market analysis on the transport landscape to discover insights, gaps and strategies. The market demand analysis will be conducted while taking into consideration various factors such as geographic, demographic, and interest to optimize strategies for CDG's Rent-a-car and chartered bus services. In addition, machine learning algorithms will be used for text analysis and market segmentation to discover insights of consumer behaviour and perception.

The dataset that we will be utilizing includes Singapore's public transport data from LTA DataMall, CDG's Facebook insights data and Power BI data. Further data collection on social media platforms and forums such as Twitter and Telegram will be conducted as well.

2.2 Motivation

CDG has several land transport business units in Singapore. However, there are still gaps in the transport market that would result in missed opportunities. With the dashboard, it can identify potential gaps in the market CDG could fill and leverage on, and better strategise market supply strategies based on various factors. Sentiments of their consumers and competitors will aid the company in further understanding areas of improvement.

2.3 Stakeholders

Sponsor	Company: ComfortDelGro Corporation Name: Gary How Relationship: Contacted him through Project Proposal Wiki Page Role: Data Team Lead
User	ComfortDelGro's Sales and Marketing Team

2.4 Deliverables

We will be delivering an engaging static web-based dashboard with the insights from our relevant respective analyses. The application will be developed using Python and PowerBI, and deployed to Microsoft Azure Cloud. This ensures that CDG's team could easily integrate with their own applications or solutions in the future. Other deliverables include source codes, documentations and or any other relevant models.

Through our deliverables, CDG will be able to better understand their consumers and market demand to accurately predict and better match consumer needs with data-driven solutions.

3.0 **Solution Architecture**: Technical Aspects

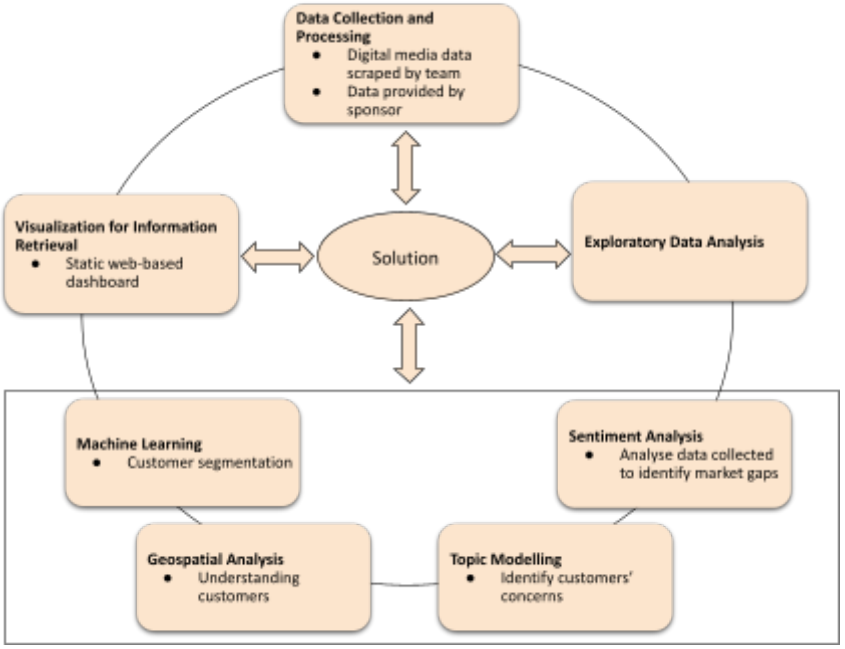


*Figure 1: **Solution Architecture** for Market Demand Analytics project*

As shown in Figure 1, we will first process the data collected via scraping of various social media platforms and cleaning of data provided by the client. Afterwards, we will conduct various analyses to identify key insights beneficial for the client, and lastly visualise the results on a dashboard for them to easily retrieve the information.

3.1 Use case 1: Data Collection and Preprocessing

The project will use data collected from Singapore's public data platform such as LTA DataMall, CDG's Facebook and PowerBI data, as well as extract data from various online media sources. The data will then be processed and cleaned and we will perform exploratory data analysis (EDA) and store them for future processing and analysis.

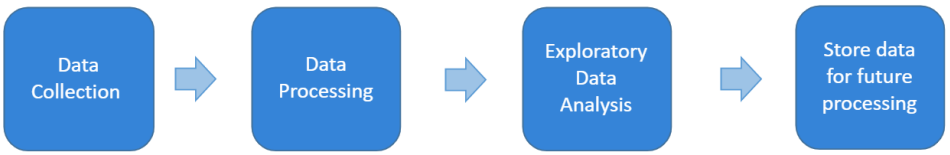


*Figure 2: **Process Flow** diagram for the solution step Data Collections and Preprocessing.*

3.2 Use Case 2: Market Demand Analysis

We will be performing text analysis on the scraped data to get customer perceptions on the overall transport landscape. Geospatial analysis and machine learning models will be performed to segment customers based on different factors such as pain points, location and interests. This project will mainly focus on optimising market supply strategies for tertiary students.

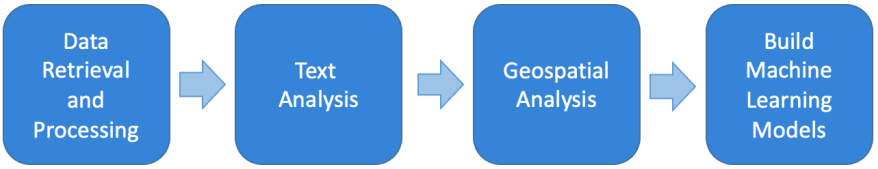


*Figure 3: **Process Flow** diagram for Analysis and Machine Learning model*

3.3 Use Case 3: Dashboard to Visualise Insights

Based on all the data insights that will be collected from our analysis, we will produce a static dashboard on Power BI to attain an overview of visualization among all the respective analyses. This would increase the company's overall efficiency and productivity.

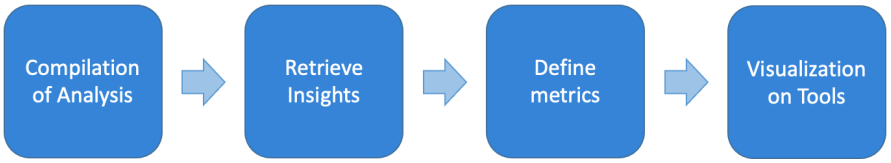


*Figure 4: **Process Flow** diagram for dashboard creation*

3.4 Features and Interactions

Please refer to Appendix A on our UML diagram. The ComfortDelGro sales and marketing team will be able to segment their customer base on certain factors such as location and interests by using the dashboard. In addition, they can view other insights driven by the relevant data.

4.0 Project Plan

4.1 Project Milestones

Please refer to Appendix B on our project timeline.

4.2 Tools and Techniques

1. Python/R for data scraping using libraries (scrapy/beautifulsoup) and used for analysis like sentiment analysis (VADER/NLTK), topic modelling, geospatial analysis (ggplot2/shiny), building machine learning model (clustering).
2. PowerBI, HTML, bootstrap for front-end development.
3. Microsoft Azure Cloud for web deployment service.

4.3 Risks

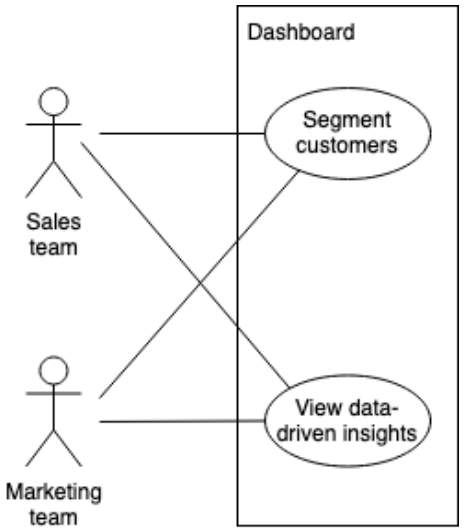
Risk	**Severity**	**Mitigation Strategy**
Data scraping and processing may take longer than expected and thus affect the timeline of the project milestones.	Likelihood: High Impact: High Risk Level: Medium	We will be working closely with the client to better understand the data they require prior to the actual start of the project. The team will also allocate buffer time for other tasks.
We might not be familiar with some of the technical skills required for the project such as PowerBI and R programming.	Likelihood: Medium Impact: High Risk Level: High	We will do our own research and learn the various technologies needed as well as approach our sponsor for any clarification.

4.4 Resource and Reference:

Collaboration: Google Drive, Microsoft Sharepoint
Project Management: Github, Monday.com
Communications: Telegram, Microsoft Teams, Whatsapp, Zoom
Resources: SMU Library, relevant documentation, DataCamp

5.0 Appendix

5.1 A: UML Diagram



5.2 B: Project Timeline