Scenario & Problem Statement:

You are a member of the data analysis team at OLIST, a Brazilian e-commerce platform. OLIST performs similar functions to Amazon by offering a marketplace solution facilitating B2B and B2C transactions. OLIST does this by supporting the entire order process including order recording, order structuring, and logistics management, and helping facilitate the buyer and seller relationship through its website.

Currently, there is uncertainty concerning how to progress OLIST's infrastructure and address late deliveries. Management wants to leverage OLIST's existing data regarding customer and seller distributions, logistics timings and seasonality, monthly order volume trends by city and state, and other relevant metrics. They wish to use this data to generate comprehensive insights to help address the concerns regarding late deliveries and to evaluate the creation of OLIST's first warehouse. Without this, management believes there is a risk of investing resources inefficiently, falling behind evolving market demands, and losing out to competition.

Purpose and Audience:

This document is intended to be used by the internal data analysis team. As such, the problem statement is viewed from a data analytics perspective. This document includes several sections relevant to the data analysis roadmap and feature creation:

1	The identification of key stakeholders
2	The identification of the potential requirements needed from various stakeholders and tools that may be used to address these requirements
3	A summary section related to each stakeholder and their requirements and how the tools may be used to meet the necessary outcome
4	A basic schema diagram of OLIST's existing data which will be used to create features to meet stakeholder requirements
5	A summary section related to each stakeholder and their requirements and which tools and how they have been used to meet the necessary outcome
6	A section which describes the limitations of the project

1. Identification of key stakeholders and responsibilities:

Executive Management	**Strategic Decision Making:** Assess and decide on the overall direction and major investments for OLIST's infrastructure and operations. **Resource Allocation:** Ensure efficient distribution of resources to address late deliveries and potential warehouse establishment. **Risk Management:** Identify and mitigate risks associated with operational inefficiencies and market competitiveness.
Business Analyst Data Team	**Data Analysis:** Extract, analyse, and interpret data related to customer and seller distributions, logistics timings, seasonality, and order volume trends. **Insight Generation:** Provide comprehensive insights and actionable recommendations based on data analysis to address late delivery issues. **Reporting:** Communicate findings and insights effectively to executive management and other stakeholders.
Operations/ Warehouse Staffing Team	**Operational Efficiency:** Implement strategies to improve order processing and delivery times. **Staff Management:** Ensure adequate staffing levels and training to meet operational demands, especially if a new warehouse is established. **Process Improvement:** Identify and recommend process improvements to enhance logistics and warehouse operations.
Supply Chain and Logistics Team	**Logistics Coordination:** Manage and optimise logistics timings and routes to minimise delivery delays. **Vendor Management:** Coordinate with suppliers and third-party logistics providers to ensure timely deliveries. **Performance Monitoring:** Track and analyse logistics performance metrics to identify and address bottlenecks or inefficiencies

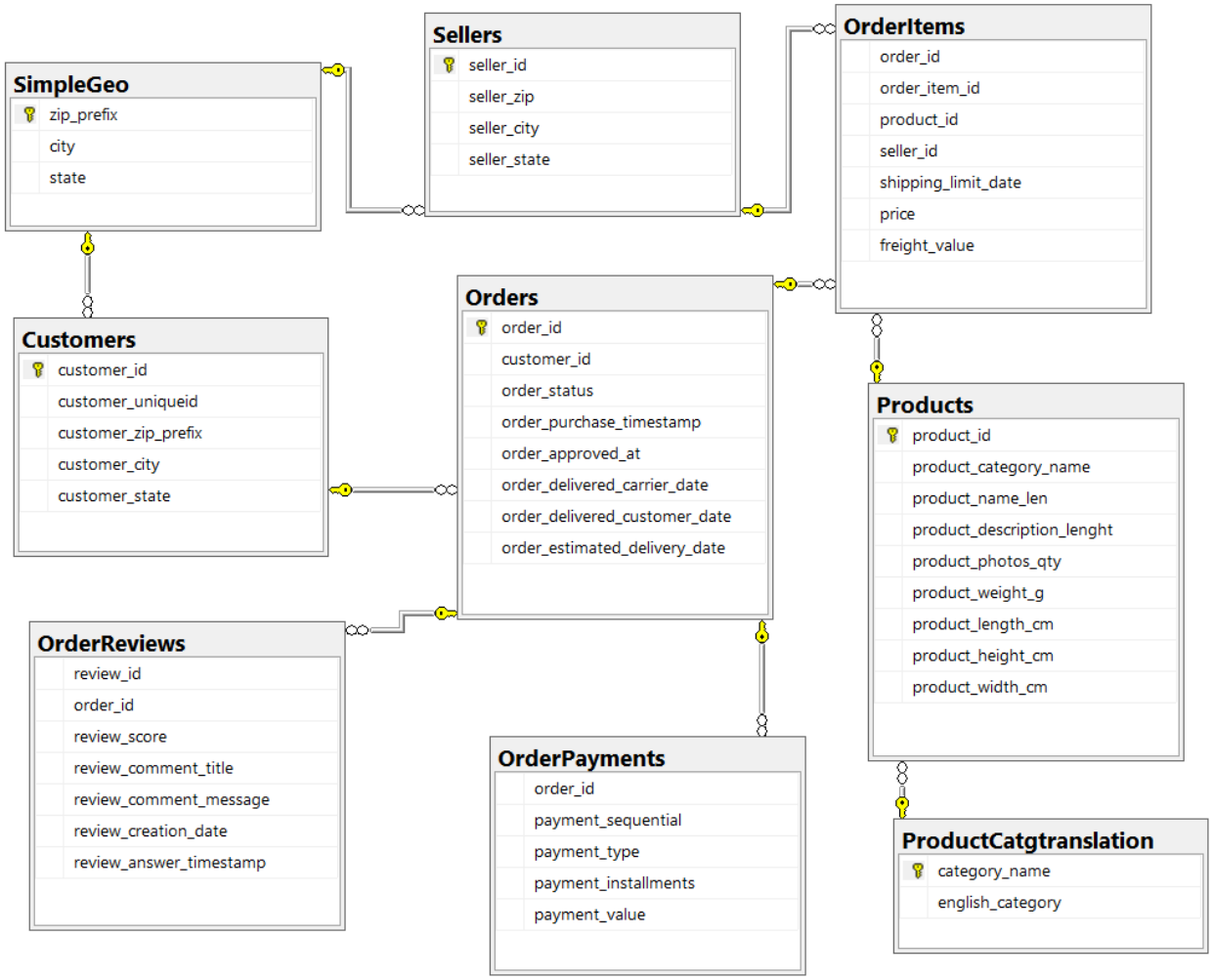
2. Identification of potential requirements from stakeholders

Stakeholder	Solution related to data (KPIs, data visuals, database querying and management)	Tool
Executive Management	A dynamic and legible report that incorporates analysis of trends regarding order volume, revenue, product categories, and revenue shares. This report should be stylistically simple and can go in-depth should the viewer wish to modify the slicers.	Power BI
Business Data Analyst Team	Transfer existing OLIST data CSVs into a local data warehouse via SQL server to enable relational database management and queries. Produce stored procedures related to Revenue, product category popularity, late order distribution, Order volume, and ranked order volumes. Using insights from queries and stored procedures, generate a comprehensive report in Power BI which satisfies Executive management's responsibilities and requirements.	SQL Server, SQL, and Power BI
Operations/Warehouse Staffing Team	Create stored procedures to retrieve monthly and daily order volumes, which can be filtered by state, city, or date. Provide a section in the report for easily viewable order volumes by state and popular product categories as well as product sizes	SQL and Power BI
Supply Chain and Logistics Team	Create stored procedures to analyse late deliveries, which can be filtered by city, state, and a specified time period. Provide a section in the report allowing for visual communication of high order volume periods and the seasonality of late delivery occurrences.	SQL and Power BI

3. How tools <u>may</u> be used to achieve the potential solution identified in section 2

Stakeholder	Tool	Tool Usage
Executive Management	Power BI	Power BI will allow the creation of interactive, dynamic reports with slicers to filter data, enabling executive management to analyse detailed order and revenue trends.
Business Data Analyst Team	SQL and Power BI	SQL will facilitate efficient data management and querying, while Power BI will visualise insights from these queries, creating comprehensive reports for management.
Operations/ Warehouse Staffing Team	SQL and Power BI	SQL will be used to develop stored procedures for detailed data retrieval, and Power BI will visualise this data, making it easy for staff to view and analyse order volumes.
Supply Chain and Logistics Team	SQL and Power BI	SQL will enable in-depth analysis of late deliveries through custom queries, and Power BI will create visual reports to highlight delivery patterns and seasonality effectively.

4. Database schema of OLIST existing data



5. How tools <u>have</u> been used to achieve proposed outcomes in section 2

Tool	Tool Usage	Outcomes achieved
SQL Server (SSMS)	-Imported existing OLIST data in the form of CSVs -Assignment of primary and foreign keys in relevant columns and between relevant fact and dimension tables -Modification of table columns to appropriate data types -Identification and removal of unverifiable rows of data (data present in fact table but not identifiable in any related dimension table) (1.5 % data loss overall)	-Creation of a relational database using OLIST data -Modified data types allow for queryable data -Removal of unidentifiable data to ensure data integrity -Allows for new data to be stored in a single space -Satisfies Data Analyst team's requirement to create a unified space where data can be stored and queried
SQL	-Queries made: - Daily, weekly, monthly order volumes by state and/or city - Matched customer-seller by state and/or city - Monthly gross, COGS, and NET revenue by state - Monthly NET revenue share by state - Summary card for late deliveries matched by same city, state, or neither - Monthly late order distributions by state - Summary card for monthly late orders - Summary card for product weight, size, and days late - Monthly popular product categories by state - Most profitable products per state - Monthly Most profitable products by state - Overall ranked product categories by product	-The created stored procedures allow the data analyst team to provide quick and valuable results when the needs arise as they are comprehensive in options and scope -Stored procedures also allow for comprehensive details to be obtained quickly even after project completion -Queried results allowed for honing in on specific insights about late deliveries order volume and state behaviour. Which, helped determine what should be included in the final report -Satisfies Operations/Warehouse Staffing Team as stored procedures allow for quick reports to be generated if need be but also order volume, product category, and order volumes by state -Satisfies Supply Chain and Logistics Team as stored procedures allow for quick reports to be generated regarding order volume for specific states or cities

	volume - Ranked product categories by state - Ranked payment methods by state -Stored procedures made: - Daily or monthly revenue by state, city, or rank range, and specified period - Daily or monthly Order volume for state, city, by rank range, and specified period - Daily or monthly popular product categories by state, city, or rank range, and specified period - Daily or monthly Order Volume by state, city (generalised, no ranks), and specified period - Daily or monthly late delivery by state, city, or rank range, and specified period	
Power BI	-Creation of a simple report which shows comprehensive insights, is stylistically simple, and easy to navigate -Report enables users to observe general insights to specific insights modified through slicers for the date and state -Customised tooltips allow for large amounts of information such as order volume, state, and net revenue -Matrix visuals were created to help visualise the relationship of order volume -Map visual to help visualise the density of order volume with aided tooltip to show order volume for states and product categories -late order distribution in the	-Stylistically simple report helps draw attention to visuals which makes them easier to read and interpret -Appropriate choice of visuals allows insights gleaned from SQL queries to be showcased in a more easily legible and interpretable manner -Dynamic dashboard through the inclusion of date and state slicers connected to multiple visuals allows users to observe general to specific insights -Satisfies Executive Management requirements as the report is easily legible but contains the most important insights. Which are related to order volume, order concentration, product popularity, and net revenue. These are all in forms that can be manipulated to showcase general to

	form of box plots to identify distribution by state and matrices to identify late deliveries by seasons	specific results depending on the user -Satisfies Operations/Warehouse Staffing Team with order volume tooltips, order volume density maps allow this stakeholder to identify key locations where staff numbers can be adjusted -Satisfies Supply Chain and Logistics team as the relationship between late orders and order volume locations sliced by time are easily identifiable

6. Limitations in regards to implementing the finalised data-driven solutions

Limitation Type	Description of Limitation
Data	Further research shows that COGS in Brazil is much more variable. Special consideration with regards to the location of a seller and customer state and/or city, as well as the product category and quantity sold all influenced the COGS. The provided data does not include any mention of this specificity. This resulted in a simplistic calculation of NET revenue to be calculated as just the product price minus its freight cost
Data	Some data had to be deleted as they were missing from key dimension tables. It is unknown whether this was by human or computer error. As the data was unverifiable by other data they had to be removed.
Data	The data set is limited and only really contains data from Jan 2017 to Aug 2018. Past Aug 2018, data uncharacteristically drops well below previous values. When observing the raw data provided by OLIST, data recording seems to begin in Jan 2017 and data collection stops after Aug 2018, thus truncating the observable data.