

Foodpanda Data Analysis

Project: Foodpanda- Customer & Order Data Analysis

Prepared For: Data Analysis Team

1.0 Project Overview

This document outlines the business requirements for analyzing Foodpanda's customer and order dataset. The primary goal is to transform raw data into actionable business intelligence to drive growth, improve customer retention, and optimize operational efficiency. The analysis will be conducted using SQL Server Management Studio (SSMS).

2.0 Business Objectives

We need to answer critical business questions to understand our performance and customer behavior. The key objectives are:

1. **Customer Analysis:** Understand our customer demographics and their value.
2. **Sales & Revenue Analysis:** Track performance and identify key revenue drivers.
3. **Customer Retention (Churn Analysis):** Identify factors leading to customer churn.
4. **Restaurant & Menu Performance:** Determine the most popular restaurants and dishes.
5. **Operational Efficiency:** Analyze delivery performance and customer satisfaction through ratings.

3.0 Detailed Requirements & Key Questions (KQs)

The analysis must provide answers to the following key questions, organized by theme.

3.1 Customer Demographics & Base Analysis

- **KQ 3.1.1:** What is the breakdown of our customer base by Gender and Age group?
- **KQ 3.1.2:** Which City has the highest number of active customers?
- **KQ 3.1.3:** What is the distribution of customers across different cities?

3.2 Sales & Revenue Performance

- **KQ 3.2.1:** What is the total revenue generated, and what is the monthly revenue trend?
- **KQ 3.2.2:** What is the average order value (AOV)?
- **KQ 3.2.3:** Which `Restaurant_Name` is the top contributor to our revenue?
- **KQ 3.2.4:** Which `Dish_Name` and `Category` are the best-selling by quantity and revenue?
- **KQ 3.2.5:** What is the most popular `Payment_Method` among customers?

3.3 Customer Behavior & Loyalty

- **KQ 3.3.1:** Who are our top 10 customers by total spending (`Price * Quantity`) and by `Loyalty_Points`?
- **KQ 3.3.2:** What is the correlation between `Order_Frequency` and total spending?
- **KQ 3.3.3:** How does the `Signup_Date` correlate with customer activity and value?

3.4 Churn Analysis (Critical)

- **KQ 3.4.1:** What is the overall customer churn rate (percentage where `Churned = 'Inactive'`)?
- **KQ 3.4.2:** What are the common characteristics (e.g., `City`, `Age`, `Order_Frequency`, average `Rating`) of churned customers?
- **KQ 3.4.3:** Is there a significant difference in the `Last_Order_Date` between active and churned customers?

3.5 Operational & Service Analysis

- **KQ 3.5.1:** What is the distribution of `Delivery_Status` (e.g., `Delivered`, `Delayed`, `Cancelled`)?
- **KQ 3.5.2:** Is there a relationship between `Delivery_Status` and the average customer `Rating`?
- **KQ 3.5.3:** What is the average `Rating` per restaurant? Identify the top and bottom 5.
- **KQ 3.5.4:** Are there specific restaurants with unusually high `Cancelled` or `Delayed` rates?

4.0 Expected Deliverables

The output of this analysis should be provided in the following formats:

1. **SQL Script File (*.sql):** A well-commented and organized script containing all the queries used for the analysis.
2. **Summary Report (Word/PDF):** A summary document explaining the findings for each Key Question, supported by key metrics and insights.
3. **Data Export (Excel/CSV):** Key results (e.g., top customers, restaurant performance) exported for further review or visualization in tools like Power BI or Excel.

5.0 Technical Specifications

- **Tool:** Microsoft SQL Server Management Studio (SSMS)
- **Data Source:** Foodpanda.csv file (to be imported into an SQL Server database table, e.g., `dbo.FoodpandaOrders`).
- **Assumptions:** The data has been cleaned and imported correctly. Date fields should be converted to the appropriate `DATE` or `DATETIME` data type during import.

6.0 Additional Notes

- Focus on clear, efficient SQL queries using aggregations (`SUM`, `COUNT`, `AVG`), `GROUP BY`, `JOINS` (if self-joining on `Customer_ID` is needed), and window functions (`RANK`, `ROW_NUMBER`) where appropriate.
- The analysis should be reproducible. The SQL script should run from beginning to end without errors.
- Highlight any data quality issues (e.g., missing values, inconsistent formatting) discovered during the analysis.