A red text on a white background

Description automatically generated

**SCHOOL OF INFORMATICS & IT**

**Storytelling Dashboard - Data Visualisation**

Student Name (Matric Number) : JAVEN LAI LE YU

Tutorial Group : P03

Jenny LING (TP)

Tutor : Nami Imanishi

Submission Date : 27/10/2023

**Declaration of Originality**

I am the originator of this work and I have appropriately acknowledged all other original sources used as my references for this work.

I understand that Plagiarism is the act of taking and using the whole or any part of another person’s work, including work generated by AI, and presenting it as my own.

I understand that Plagiarism is an academic offence

and if I am found to have committed or abetted the offence of plagiarism in relation to this submitted work, disciplinary action will be enforced.

**Declaration on the use of Generative AI tools for assignments**

|  |
| --- |
| Describe how you have used Generative AI tools such as ChatGPT or Dall.E-2 in your assignment.  Show snapshots of the conversations with the AI tool (i.e., the prompts you used and the response you get from the AI tool). |
| For purpose of Report, I used ChatGPT to clarify the project requirements stated in the specifications and used it to generate content for my report. Additionally, I used it to rewrite my explanations to sound more fluent, clear, and professional. Lastly, I also used it to clarify some math logical calculation when listing out statistical figures at task 3.  For purpose of Tableau and Alteryx, I asked ChatGPT to construct the formula to perform data transformation based on my given description. The formula was constructed with ChatGPT, and not taken completely off ChatGPT, as the AI wasn’t able to always produce a formula that works exactly as I want, and I had to adjust.  E.g. Asking ChatGPT to put columns into list, to save time from copying one by one:    Asking ChatGPT to make explanations more fluent:    Asking ChatGPT to rephrase a word with something more appropriate:    Asking ChatGPT to calulate by putting ? in the value:    ChatGPT to assist with creation of formula to create new calculated field:    Read full conversation here:  <https://chat.openai.com/share/f8ec7450-5b3d-4691-adb1-255d248c6a46> |
| How do you indicate the reference?  The content generated by AI tools are not retrievable except by the user who generated them, so they are considered non-recoverable sources. Although non-recoverable data or quotations in APA Style papers are usually cited as personal communications, with ChatGPT-generated text there is no person communicating. Quoting text from ChatGPT chat is therefore more like sharing the output of an algorithm, with a reference list entry and the corresponding in-text citation.  According to the official APA Style site, ChatGPT references should be cited as:  E.g. OpenAI. (2023). *ChatGPT* (Sep 25 version) [Large language model].  <https://chat.openai.com/chat> |

**Content Page**

**Task 1: Background and Business Requirements**

* 1. Background ------------------------------------------------------------------------------------------- 6
  2. Problem ----------------------------------------------------------------------------------------------------- 6
  3. Purpose ----------------------------------------------------------------------------------------------------- 6
  4. Objective ----------------------------------------------------------------------------------------------------- 6

**Task 2: Data Management**

2.1 Data Profiling ----------------------------------------------------------------------------------------- 6

2.1.1 Data Attributes -------------------------------------------------------------------------------------- 6 – 7

2.1.2 Datasets Analysis ------------------------------------------------------------------------------------ 7 – 8

2.1.3 Data Quality Analysis ------------------------------------------------------------------------------- 8 – 9

2.2 Data Integration ----------------------------------------------------------------------------------------- 10

2.3 Features Creation ---------------------------------------------------------------------------------------- 11

**Task 3: Data Exploration**

3.1 Customer Satisfaction Analysis -------------------------------------------------------------------- 12

3.2 Common Reasons for Inquiry Analysis ------------------------------------------------------------ 12

3.3 OCBC Branches Analysis ----------------------------------------------------------------------------- 13

3.4 Demand of Inquiries throughout the day Analysis --------------------------------------------- 13

**1. Background and Business Requirement**

1.1 **Background**: Overseas-Chinese Banking Corporation (OCBC) primarily relies on physical walk-in inquiries at our branches and phone calls for customer inquiries and complaints. However, these channels are often **overwhelmed by the volume of inquiries from customers**, and OCBC **customer service** **agents are struggling to keep up** with the demandand respond to all the inquiries. This not only **impacts operational efficiency**, but also leads to **inconsistent customer experience** across different branches and customers.

1.2 **Problem**: This inconsistency in customer experience is a critical concern as it has the potential to **erode customer trust and loyalty**. In the highly competitive banking industry, where customers have a plethora of financial institutions to choose from, **poor customer service can drive them to seek services from competitors**. As such, there is an **urgent need to evaluate and enhance customer satisfaction** at OCBC, hence the need for this analysis.

1.3 **Purpose**: This analysis is driven by the imperative to **understand customer satisfaction** within OCBC. It endeavours to **understand and offer valuable insights about why our customers are reaching out** to us, along with actionable strategies to manage the overwhelming demand for OCBC helpdesk and support OCBC in its commitment to delivering a superior banking experience to every customer.

1.4 **Objective**: The objectives of this analysis are stated below.

1.4.1 Find out if OCBC customers are currently satisfied with our customer service.

1.4.2 Find out the most common reason(s) for calls or branch visits.

1.4.3 Identify factors that affect feedback score and explain how they influence the score.

1.4.4 Identify peak time periods when both channels experience the highest activity.

1.4.5 Identify if there is a specific time when a channel is preferred over the other and why.

**2. Data Management**

**2.1 Data Profiling**

The datasets provided by OCBC Database Administrators include the bank’s records of customer service agent engaged in phone calls and walk-in inquiries from 2021 and 2022. Additionally, a file containing the location of the branches is provided.

**2.1.1 Data Attributes**

**Call Log**

|  |  |  |  |
| --- | --- | --- | --- |
| Field Name | Description | Measurement | Format |
| Customer ID | Unique customer ID for call log | Nominal | C10001 |
| Date | Date of the call | Interval | dd/MM/yyyy |
| Call Start Date and Time | When the agent begins engaging with the customer on the call | Interval | dd/MM/yyyy hh:mm |
| Call End Date and Time | When the agent ends the call | Interval | dd/MM/yyyy hh:mm |
| Phone Line Agent ID | Unique identifier for an agent | Nominal | A1001 |
| Reason for Call | The reason why the customer contacted OCBC | Nominal | self service inb - enquiry |
| Feedback | Rating of the call from 1 to 5, with 5 being the highest rating | Ordinal | 1 |

**Walk-in**

|  |  |  |  |
| --- | --- | --- | --- |
| Field Name | Description | Measurement | Format |
| Branch ID | Branch location ID | Nominal | B001 |
| Branch Area | Location of the Branch | Nominal | North |
| Customer ID | Unique customer ID for walk-in | Nominal | W123456 |
| Customer Age | Customer Age range | Interval | 41-50 |
| Customer Gender | Gender of Female & Male (F/M) | Nominal | F |
| Date | Date of the walk-in to the branch | Interval | dd/MM/yyyy |
| Waiting time (min) | Duration of the waiting time in minutes | Ratio | 23 |
| Start Date and Time | When the agent starts serving the customer | Interval | dd/MM/yyyy hh:mm |
| End Date and Time | When the agent ends the service | Interval | dd/MM/yyyy hh:mm |
| Agent ID | Unique identifier for an agent | Nominal | A0011001 |
| Reason for visit | Purpose behind why customers visit the OCBC Branch | Nominal | Balance transaction enquiry |
| Customer Satisfaction Rate | Customer satisfaction rating from 1 to 5, with 5 being the highest rating | Ordinal | 1 |

**Branch Location Singapore**

|  |  |  |  |
| --- | --- | --- | --- |
| Field Name | Description | Measurement | Format |
| Branch ID | Unique Branch location ID | Nominal | B001 |
| Latitude | North-South position | Nominal | 1.361562 |
| Longitude | East-West position | Nominal | 103.8415 |

A screenshot of a computer screen

Description automatically generated**2.1.2 Datasets Analysis**

Inspecting the Branch Location Singapore dataset using Excel, **there are 15 OCBC Branches**, and the coordinates have no data quality issue.

Alteryx Designer (ETL Software) will be used for data inspection and cleaning for the walk-in and call log datasets as those datasets are too large to manually inspect using Excel.

The Field Summary Node will be used to inspect and audit each dataset for potential data quality issues to be found and fixed.

[Screenshot of Branch Location Singapore on Excel]

The **datasets for 2021 and 2022 were received separately** for both call log and walk-in. Both **call log and walk-in datasets had similar fields**: Customer ID, Date, Start Time, End Time, Agent ID, reason for inquiry, and rating score out of 5.

The walk-in dataset has additional fields like branch ID, branch area, and waiting time. The **Agent ID for walk-in and call logs varies in format** which indicates that the customer service agents are not related. Walk-in datasets have Branch ID with the same format as the Branch Location Singapore table, hence **Branch ID can be the common key to link and extract the coordinates** of the branch. However, **call logs have no Branch ID** field and **cannot be linked with an OCBC branch**.

**Call-Log 2021:**

A screenshot of a computer

Description automatically generated

[Missingness percentage of Feedback, Missing ‘C’ at Customer ID records]

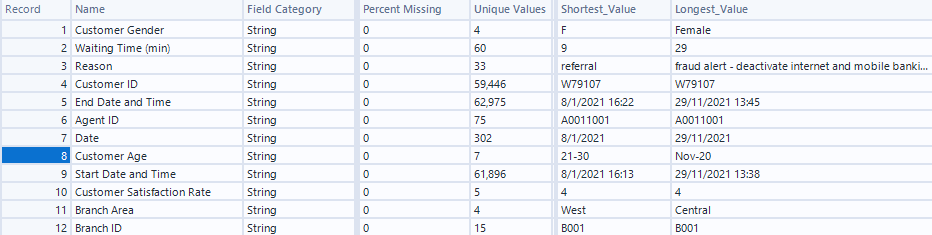
**Call-Log 2022:**

A screenshot of a computer

Description automatically generated

[Missingness percentage of Feedback, Inconsistent date format]

**Walk-in 2021:**



[Unstandardized customer genders, customer age range auto-formatted into a date]

**Walk-in 2022:**

A screenshot of a computer

Description automatically generated

[column names differ from 2021 dataset, customer age range auto-formatted into a date, Problem with some Branch IDs]

**2.1.3 Data Quality Analysis**

**Accuracy:** The **datasets are not in time-series** **sequential order**. This makes the reliability and accuracy of the datasets questionable as the records should be automatically logged into the system after each call or walk-in inquiry ticket in a systematic order.

*Refer to the photo below.*

A screenshot of a table

Description automatically generated

**Timeliness:** The datasets are recent as they are from the past 2 years.

**Consistency:**

1. Call-log 2021 has inconsistency in **Customer ID where some IDs were missing 'C'** infront.
2. Call-log 2022 has **inconsistent date format** **in Date** where values to do follow the dd/MM/yyyy format.
3. Gender label for customers in walk-in 2021 records is inconsistent as there are values **Female/Male when it should be F/M**.
4. A graph of blue bars

   Description automatically generated with medium confidenceCustomer Age range has **mislabelled range of Nov-20 instead of 11-20**. I discovered that Nov-20 was supposed to represent 11-20 as Nov represents 11 on the calendar. Therefore, it's justified to assume that the system misunderstood 11-20 and autoformatted into a date.
5. Column Start Date and Time and column End Date and Time have **different column names** for walk-in 2022. I need to standardize the column names in order to union walk-in 2021 with 2022 together.
6. A graph with blue dots and numbers

   Description automatically generatedWalk-in 2022 has inconsistencies with Branch ID as there are **Branch IDs that are missing the B00 header**. This can be confirmed as these branch numbers have the same branch area as the branch ID where ID = B000 + number.



**Completeness:**

1. There are **30% of Feedback scores for call-logs are missing**. Using context knowledge, these values are **MNAR** as it is likely that the customers refused to leave feedback after the call as they couldn’t be bothered. Hence, **no imputation is required as the missing value has its own meaning**.
2. **Call logs are missing fields like waiting time**, customer age, and Branch ID, deeming call logs to be less useful than walk-in as there are fewer fields to work with.

**Relevance:** As the datasets contain a rating score that represents customer satisfaction, reason for inquiry, waiting time, customer age, and other factors about OCBC inquiry customers, the datasets provided are highly relevant for my analysis in understanding why customers are reaching out to OCBC.

**2.2 Data Integration**

2.2.1 Union of datasets: After performing the necessary data cleaning to address the data issues stipulated in data profiling using Alteryx, I created the cleaned datasets into new CSV files. Then, I **union the 2021 and 2022 records for walk-in and call log** respectively, **as the columns for 2021 are the same for 2022** and **2022 is a continuation** of 2021 records for both walk-in and call log.

A screenshot of a computer

Description automatically generated A screenshot of a computer

Description automatically generated

[Snapshots of how the datasets are union together using Tableau (Software)]

2.2.2 Relationship connection: A relationship between the 2021 union 2022 walk-in dataset and the Branch Location dataset is created, using Branch ID as a common key to link these 2 datasets together. This relationship allows the longitude and latitude of a branch to be accessed from the walk-in dataset, to enable the creation of map graphs. [Relationships are more flexible and dynamic than Joins](https://help.tableau.com/current/online/en-us/datasource_relationships_learnmorepage.htm#:~:text=Relationships%20are%20a%20dynamic%2C%20flexible%20way%20to%20combine,Use%20joins%20only%20when%20you%20absolutely%20need%20to.), which is why it is chosen over Joins.

A screenshot of a computer

Description automatically generated

[Snapshot of relationship link between the 2 datasets using Tableau]

2.2.3 Organization of data source: The combined call dataset and combined walk-in dataset + branch location dataset are stored in different data sources as the requirements indicate to analyse call and walk-in separately.

**2.3 Features Creation**

2.3.1 Enquiry:

This field **summarizes the reason for the customer reaching out to OCBC in a maximum of 2 words**. This is useful as some Reason descriptions are very long, making it unsightly and difficult to read on a visualization, hence a brief and concise description would be better. This column will be used to identify the main reasons for call-in and walk-in inquiries.

Formula:

SPLIT(REPLACE([Reason], "\_", " "), " ", 1)

+ " " +

SPLIT(REPLACE([Reason], "\_", " "), " ", 2)

* + 1. Duration:

This field represents **how long the inquiry session lasted, in minutes**. It is useful for understanding how long our customers consult OCBC for an inquiry. It can be added to tooltips or labels in a graph, contributing as an additional information which may assist with the derivative of insights.

Formula:

DATEDIFF('minute', [Start Time], [End Time])

* + 1. Number of Agents present:

This field represents **how many customer service agents were present during a particular hour of a day**. It is useful for assessing the supply of customer service support to find out if OCBC is short-staffed, and by how much. It can be used as a measure to be compared with other measurements to create a graph.

Formula:

{ FIXED DATETRUNC('hour', [Call Start]): COUNTD([Phone Line Agent ID]) }

* + 1. No Feedback [Set for NULL values in Feedback]:

This field contains **records with missing values for Feedback**. It is required for segregating missing values to prevent these records from interfering with or tainting the analysis of customer satisfaction. It can be used to differentiate records using filters.

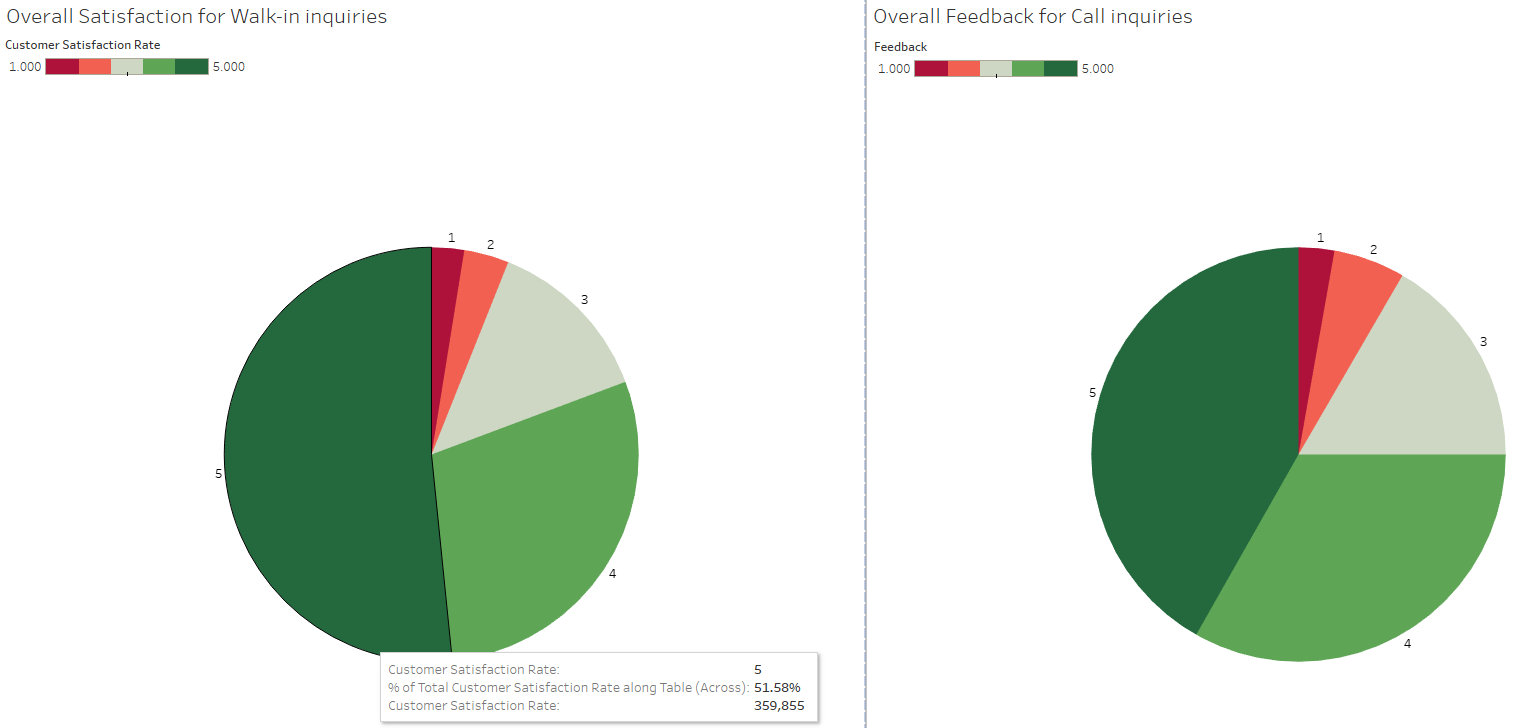
Configuration:

A screenshot of a computer screen

Description automatically generated

**3. Data Exploration**

**3.1 Customer Satisfaction Analysis:**

****

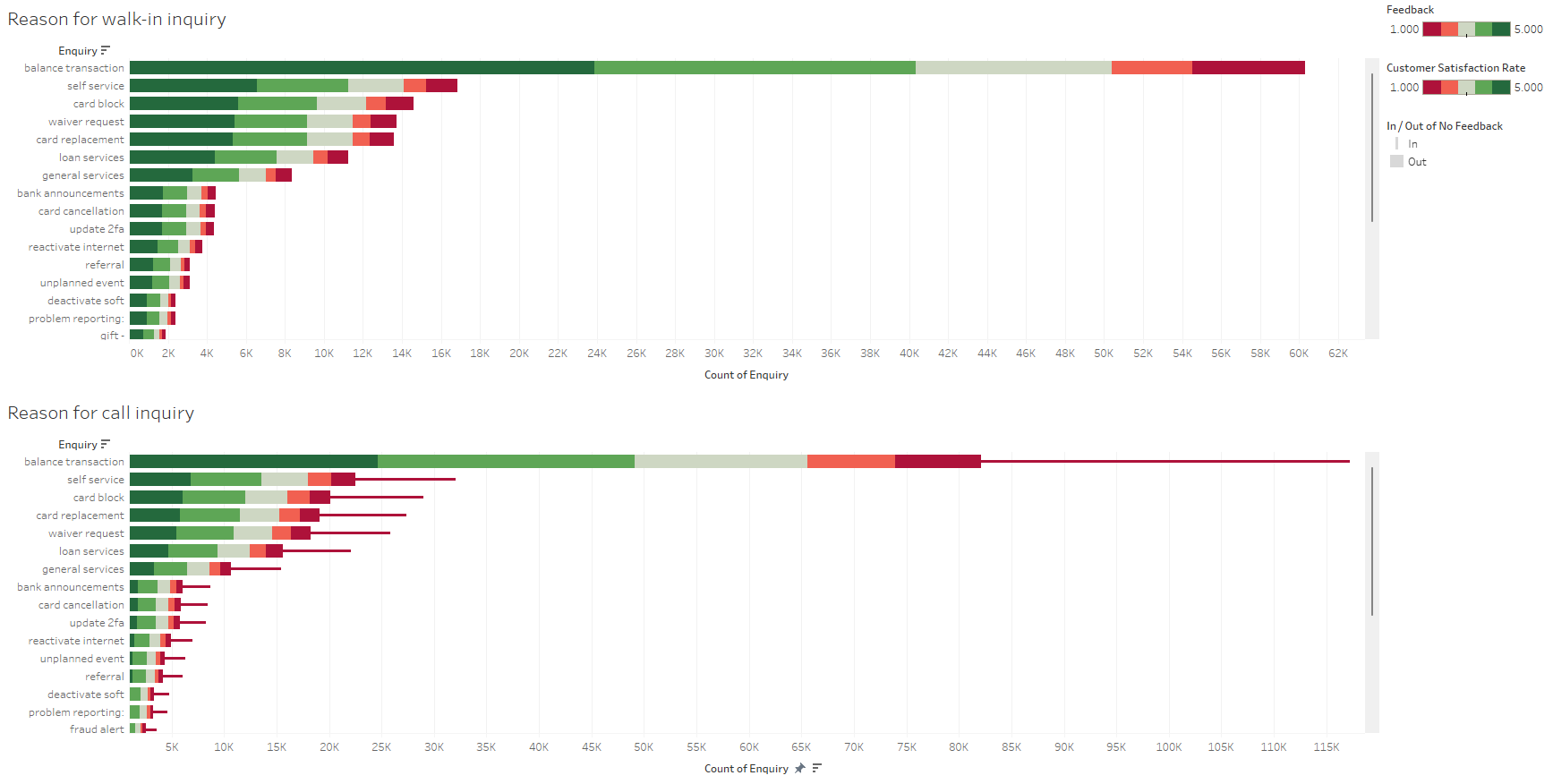
3.1.1 *Null Feedback records are not included in this analysis.*

3.1.2 The pie charts on overall satisfaction indicates a **commendable performance by the customer service agents**. Majority of customers are satisfied as they give a rating of 3 - 5.

3.1.3 **Over 50% of customers who engaged with OCBC's walk-in customer service were completely satisfied** and gave a 5/5 rating.

3.1.4 The performance for call-ins is slightly poorer than walk-in inquiries as the percentage of 4-5/5 is smaller, which may put the **call-in customer service’s pristine reputation at risk**.

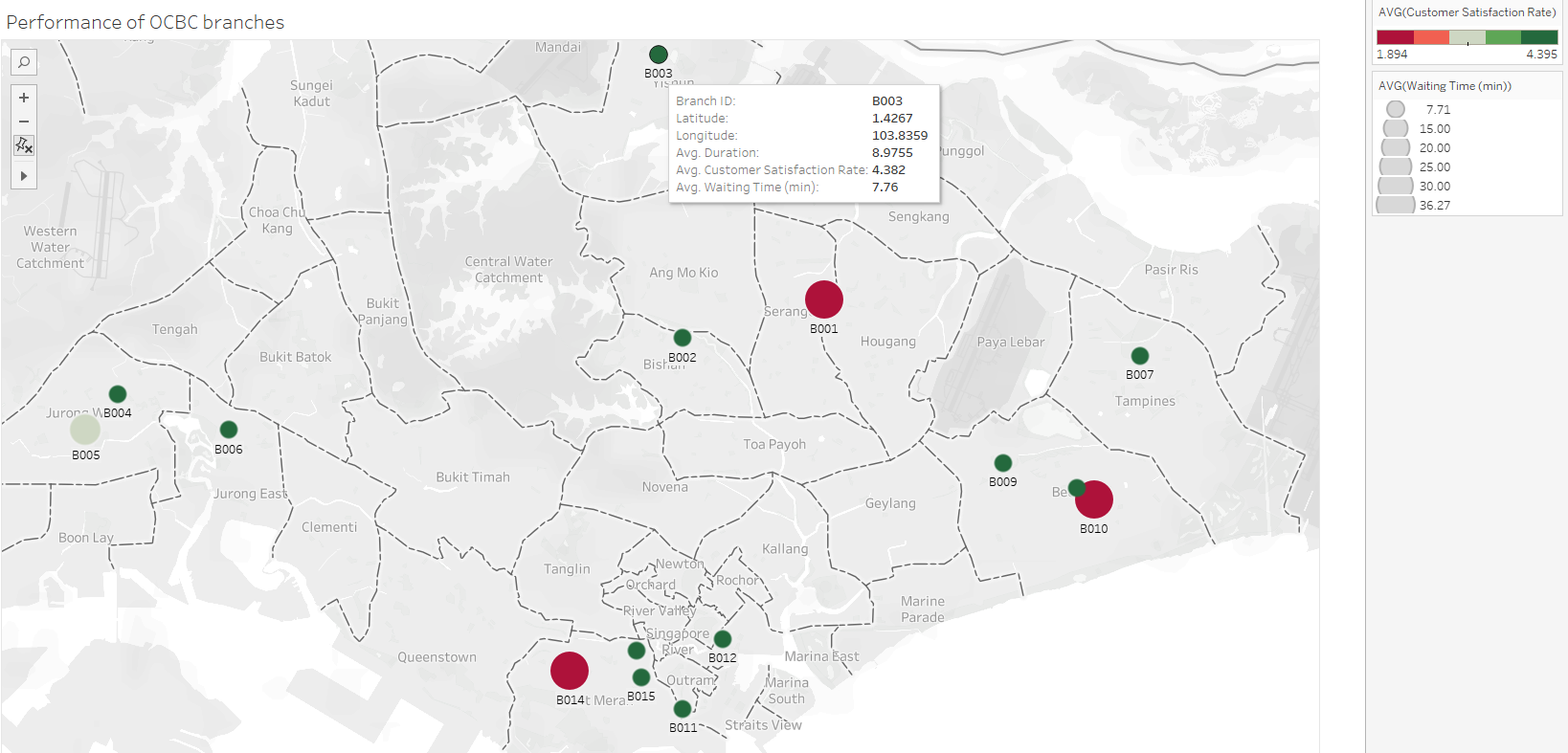
**3.2 Common Reasons for Inquiry Analysis:**

**

3.2.1 The bar chart displays the frequency of inquiry types, ordered by popularity. The **most common inquiries are related to bank transactions**, followed by services and card-related inquiries.

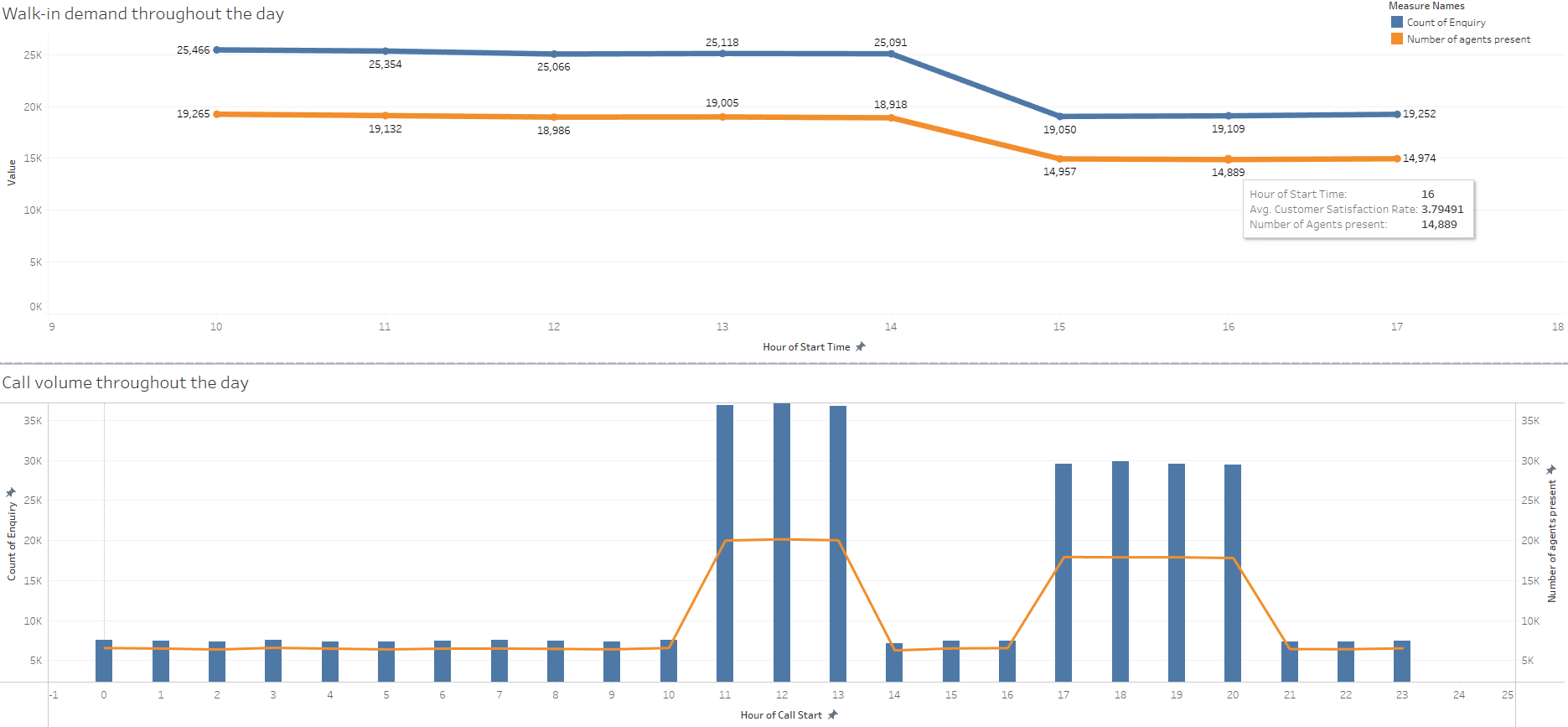
3.2.2 Breaking each inquiry down by satisfaction rate, there is no particular inquiry that has a prevalent bad satisfaction rate. This implies that OCBC’s **customer service team is adept at handling various inquiries satisfactorily**. The thinner bar for the call-in graph represents customers who did not leave feedback and does not represent negative ratings.

**3.3 OCBC Branches Analysis:**



3.3.1 Examining individual branches, a **clear factor affecting unsatisfactory performance appears to be substantially longer waiting times**. Branches with poor satisfaction ratings average significantly extended waiting times. Notably, this relationship doesn't appear to be influenced by geographical location. For instance, Bedok has multiple branches, but only the branch with substantially longer waiting times experiences poorer average ratings. This implies a direct **inverse relationship between customer satisfaction and waiting time**.

**3.4 Demand of Inquiries throughout the day Analysis:**



3.4.1 **Peak inquiry hours** for both walk-in and call inquiries are from **11 AM to 1 PM**.

3.4.2 There is **high service demand from 5 PM to 8 PM**, but inquiries can **only be received through call-ins** during this time as OCBC branches stop taking walk-in inquiries after 5 PM.

3.4.3 There's a clear **shortage of agents during call-in peak hours**, indicated by a supply-to-demand ratio of 0.5 to 0.6. This means each staff member has to attend to nearly 2 customers concurrently, which may adversely affect their performance. Hence, there is a **need for additional resources during these high-demand periods** to ensure outstanding customer service standards at all times.

End of Part 1