

# **SCHOOL OF INFORMATICS & IT**

# **Storytelling Dashboard - Data Visualisation**

Student Name (Matric Number) : JAVEN LAI LE YU

Tutorial Group : P03

Tutor : Nami Imanishi

Submission Date : 31/10/2023

# **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**

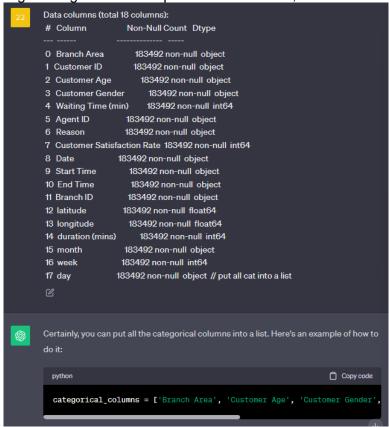
Describe how you have used Generative AI tools such as ChatGPT or Dall.E-2 in your assignment.

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 Al 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:



2. Branch Performance Impact: Examining individual branches, a clear factor affecting unsatisfactory performance appears to be longer waiting times. Branches with poor satisfaction ratings are correlated with significantly extended waiting times. It is also proven that it is not due to geographical location as Bedok has multiple branches but only the branch with substantially longer waiting time has a poor average rating. This suggests a direct relationship between customer satisfaction and waiting time. // is there a way to rephrase "proven"





1. Branch Performance Impact: When scrutinizing individual branches, a distinct factor contributing to unsatisfactory performance becomes evident: longer waiting times. Branches with lower satisfaction ratings exhibit a strong correlation with considerably prolonged waiting times. Notably, this relationship doesn't appear to be influenced by geographical location. For instance, in Bedok, where multiple branches are present, only the branch with significantly extended waiting times experiences poorer average ratings. This strongly implies a direct link between customer satisfaction and waiting time.

# Asking ChatGPT to rephrase a word with something more appropriate:

22

4. 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 hindering the analysis.

// replace hindering with an appropriate word that means polluting/contaminating the reseult

[4



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 tainting the analysis.

<u>a</u> a s

# Asking ChatGPT to calulate by putting? in the value:

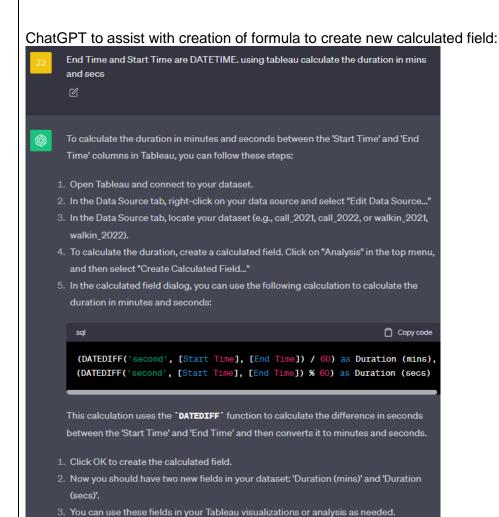


4. Analysing the demand for inquiries throughout the day, it is evident that demand is consistently high from 10 AM (opening hour) till 2 PM. Demand for inquiries is still present from 3 PM till the closing hour of 5 PM. The amount of helpdesk staff available for each working hour closely follows the trend line of the customer service demand, with a supply-to-demand ratio of 0.76-0.78. This means that 1 helpdesk staff has to attend to ? customer inquiry for an hour on average.





1. Analyzing the demand for inquiries throughout the day, it is evident that demand is consistently high from 10 AM (opening hour) till 2 PM. Demand for inquiries is still present from 3 PM till the closing hour of 5 PM. The number of helpdesk staff available for each working hour closely follows the trend line of the customer service demand, with a supply-to-demand ratio of 0.76-0.78. This means that, on average, 1 helpdesk staff has to attend to approximately 1.28 to 1.32 customer inquiries per hour.



# Read full conversation here:

https://chat.openai.com/share/f8ec7450-5b3d-4691-adb1-255d248c6a46

By following these steps, you can calculate the duration in minutes and seconds

# 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. OpenAl. (2023). *ChatGPT* (Sep 25 version) [Large language model]. https://chat.openai.com/chat

# **Content Page**

Task 1: Background and Business Requirements	
1.1 About Report	6
1.2 Background	6
1.3 Problem	6
1.4 Purpose	
1.5 Stakeholders	
1.6 Objective	6
Task 2: Data Management	
2.1 Data Profiling	
2.1.1 Data Attributes	7
2.1.2 Datasets Analysis	
2.1.3 Data Quality Analysis	9 <b>-</b> 10
2.1.4 Data Cleaning Summary	
2.2 Data Integration	11
2.3 Features Creation	12
Task 3: Data Exploration	
3.1 Customer Satisfaction Analysis	
3.2 Common Reasons for Inquiry Analysis	
3.3 OCBC Branches Analysis	14
3.4 Demand of Inquiries throughout the day Analysis	14

# 1. Background and Business Requirements

- 1.1 **About Report**: This report is an evaluation of the Customer Service experience at OCBC, to be distributed to Customer Service Executives and Customer Service Managers at OCBC.
- 1.2 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 demand and respond to all the inquiries. This not only impacts operational efficiency, but also leads to inconsistent customer experience across different branches and customers.
- 1.3 **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.
- 1.4 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.

## 1.5 Stakeholders:

- 1.5.1 **Customer Service Executives/Managers**: Your role is to consider actions to be taken based on the insights from this report and disseminate pertinent information to subordinates.
- 1.5.2 **Customer Service Agents**: The insights and recommendations should be shared with our agents to provide them with guidance on areas for improvement. Customer Service Agents are the direct recipients of this project.
- 1.5.3 **Customers**: This project aims to improve their experience with OCBC customer service. Hence, they are the direct benefactors of this project.
- 1.6 **Objective**: The objectives of this analysis are stated below.
- 1.6.1 Find out if OCBC customers are currently satisfied with our customer service.
- 1.6.2 Find out the most common reason(s) for calls or branch visits.
- 1.6.3 Identify factors that affect feedback score and explain how they influence the score.
- 1.6.4 Identify peak time periods when both call-in and walk-in channels experience the highest activity.
- 1.6.5 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	Duration of the waiting time in minutes	Ratio	23		
(min)					
Start Date and	When the agent starts serving the	Interval	dd/MM/yyyy		
Time	customer		hh:mm		
End Date and	When the agent ends the service	Interval	dd/MM/yyyy		
Time			hh:mm		
Agent ID	Unique identifier for an agent	Nominal	A0011001		
Reason for visit	Purpose behind why customers visit	ny customers visit Nominal			
	the OCBC Branch		enquiry		
Customer	Customer satisfaction rating from 1 to	Ordinal	1		
Satisfaction Rate	5, with 5 being the highest rating				

# **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

# 2.1.2 Datasets Analysis

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

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

latitude	longitude
1.370405	103.874
1.361562	103.8415
1.426725	103.8359
1.348676	103.7116
1.340475	103.7043
1.340486	103.7371
1.357508	103.9466
1.327163	103.9321
1.332822	103.9152
1.32456	103.936
1.276376	103.8416
1.292343	103.8507
1.289681	103.831
1.28517	103.8156
1.283662	103.832
	1.370405 1.361562 1.426725 1.348676 1.340475 1.340486 1.357508 1.327163 1.332822 1.32456 1.276376

[Screenshot of Branch Location Singapore on Excel]

- D) The datasets for 2021 and 2022 were received separately for both call log and walk-in. Both call log and walk-in datasets have similar fields: Customer ID, Date, Start Time, End Time, Agent ID, reason for inquiry, and a rating score out of 5.
- E) 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**.
- F) Call-log and Walk-in datasets have a common key, Date, that could link the datasets. However, because the specifications requires me to analyse walk-in and call-logs separately and it is unmeaningful to link these 2 datasets, I will not consider joining these 2 datasets.

#### **Call-Log 2021:**

	0						
Record	Name	Field Category	Percent Missing	Unique Values	Remarks	Shortest_Value	Longest_Value
1	Phone Line Agent ID	String	0	50	Some values of this field	A1000	A1000
2	Customer ID	String	0	17,124	Some values of this field	10099	C10003
3	Call End Date and Time	String	0	121,610	Some values of this field	1/1/2021 8:19	10/10/2021 10:39
4	Call Start Date and Time	String	0	109,103	Some values of this field	1/1/2021 8:08	10/10/2021 10:27
5	Reason for Call	String	0	33	Some values of this field	referral	fraud alert - deactivate internet and mobile l
6	Date	String	0	336	Some values of this field	1/1/2021	10/10/2021
7	Feedback	String	30.002901	6	This field has over 10% m	1	1

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

## **Call-Log 2022:**

Record	Name	Field Category	Percent Missing	Unique Va	Remarks	Shortest_Value	Longest_Value
1	Phone Line Agent ID	String	0	50	Some values of this field have a small n	A1019	A1019
2	Customer ID	String	0	17,081	Some values of this field have a small n	C10000	C10000
3	Call End Date and Time	String	0	120,117	Some values of this field have a small n	1/1/2022 0:13	10/10/2022 10:03
4	Call Start Date and Time	String	0	107,572	Some values of this field have a small n	1/1/2022 0:03	10/10/2022 10:03
5	Reason for Call	String	0	33	Some values of this field have a small n	referral	fraud alert - deactivate internet and mobile banki
6	Date	String	0	383	Some values of this field have a small n	1-Jan-22	10/10/2022
7	Feedback	String	30.002557	6	This field has over 10% missing values	5	5

[Missingness percentage of Feedback, Inconsistent date format]

#### Walk-in 2021:

Record	Name	Field Category	Percent Missing	Unique Values	Shortest_Value	Longest_Value
1	Customer Gender	String	0	4	F	Female
2	Waiting Time (min)	String	0	60	9	29
3	Reason	String	0	33	referral	fraud alert - deactivate internet and mobile banki
4	Customer ID	String	0	59,446	W79107	W79107
9	End Date and Time	String	0	62,975	8/1/2021 16:22	29/11/2021 13:45
6	Agent ID	String	0	75	A0011001	A0011001
7	Date	String	0	302	8/1/2021	29/11/2021
8	Customer Age	String	0	7	21-30	Nov-20
9	Start Date and Time	String	0	61,896	8/1/2021 16:13	29/11/2021 13:38
10	Customer Satisfaction Rate	String	0	5	4	4
11	Branch Area	String	0	4	West	Central
12	Branch ID	String	0	15	B001	B001

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

#### Walk-in 2022:

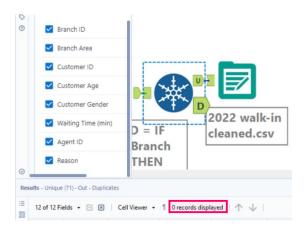
Record	Name	Field Category	Percent Missing	Unique Values	Remarks	Shortest_Value	Longest_Value
1	End Time	String	0	65,031	Some values	2/5/2022 16:51	11/11/2022 14:02
2	Waiting Time (min)	String	0	32	Some values	7	43
3	Reason	String	0	33	Some values	referral	fraud alert - deactivate internet and
4	Customer ID	String	0	60,699	Some values	W00000	W00000
5	Agent ID	String	0	75	Some values	A0141000	A0141000
6	Date	String	0	313	Some values	2/5/2022	11/11/2022
7	Start Time	String	0	64,059	Some values	2/5/2022 16:40	11/11/2022 13:47
8	Customer Age	String	0	7	Some values	31-40	Nov-20
9	Branch ID	String	0	24	Some values	6	B014
10	Customer Gender	String	0	2	[Null]	F	F
11	Customer Satisfaction Rate	String	0	5	[Null]	2	2
12	Branch Area	String	0	4	[Null]	East	Central

[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

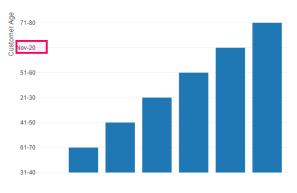
Accuracy: The datasets are not in time-series sequential order. This makes the reliability and accuracy of the datasets questionable as the records should be automatically logged into the system after each call or walk-in inquiry ticket in a systematic order.

All datasets have no duplicated rows based on Unique Node in Alteryx. This indicates some reliability in the datasets.



#### **Consistency:**

- Call-log 2021 has inconsistency in Customer ID where some IDs were missing 'C' infront. Call-log 2022 has inconsistent date format in Date where some values do not follow the dd/MM/yyyy format as per data dictionary. This column has to be standardized to ensure all values comply with the format as per data understanding.
- 2. Gender label for customers in walk-in 2021 records is inconsistent as there are values **Female/Male when it should be F/M**. I will **standardize this column** to only contain F or M as per data dictionary understanding.
- 3. Customer 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 and there is no age range for youths below 21. Therefore, it's justified to assume that the system misunderstood 11-20 and autoformatted into a date.



4. 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 to ensure all columns are the same for 2021 and 2022, in order to union walk-in 2021 with 2022.

5. Walk-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.



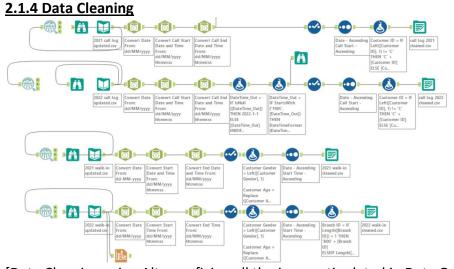
I must append the B00 header for Branch IDs with a missing header, to ensure all walk-in records can successfully link with their respective Branch ID in the Branch Location dataset.

#### **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, 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.

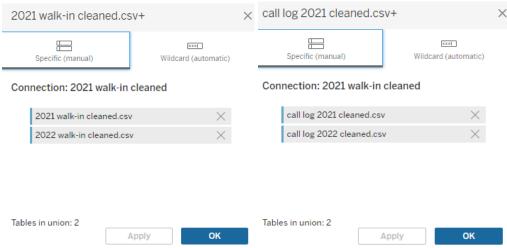
**Timeliness:** The datasets are recent as they are from the past 2 years, hence they should reflect the current trend and status of customer service satisfaction in OCBC.



[Data Cleaning using Alteryx, fixing all the issues stipulated in Data Quality Analysis]

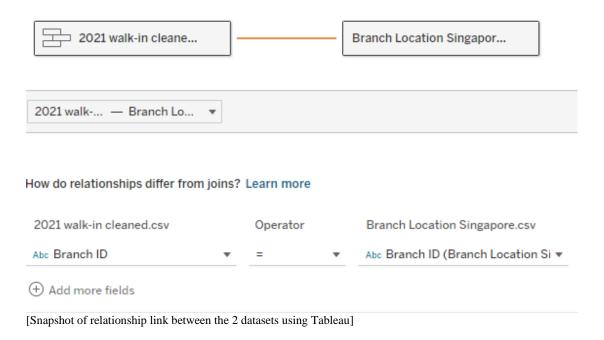
# 2.2 Data Integration

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.



[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, which is why it is chosen over Joins.



2.2.3 **Organization of data source**: The combined call dataset (1) and combined walk-in dataset + branch location dataset (2) are stored in 2 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)
```

#### 2.3.2 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])
```

# 2.3.3 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.

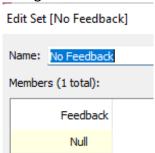
#### Formula:

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

## 2.3.4 No Feedback [Set for NULL values in Feedback]:

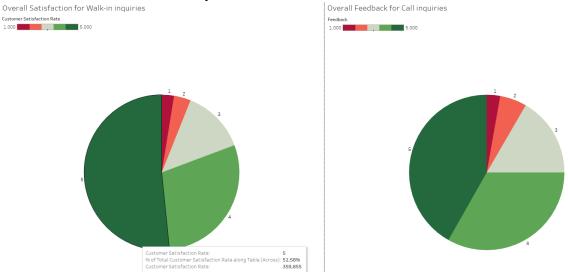
This field contains **records with missing values for Feedback**. It is required for segregating missing values to prevent these records from interfering with or tainting the analysis of customer satisfaction. It can be used to differentiate records using filters.

## Configuration:



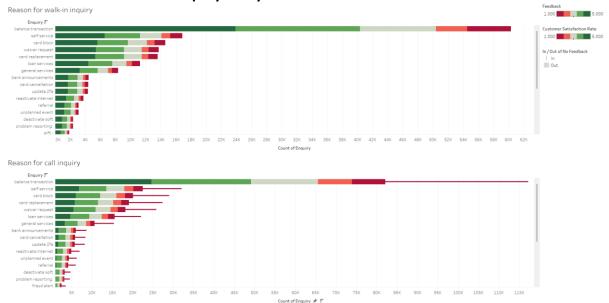
# 3. Data Exploration

## 3.1 Customer Satisfaction Analysis:



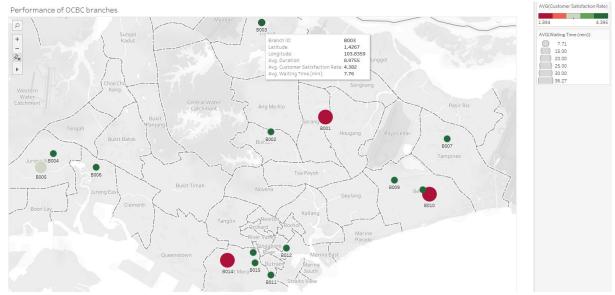
- 3.1.1 Null Feedback records are not included in this analysis.
- 3.1.2 The pie charts on overall satisfaction indicates a **commendable performance by the customer service agents**. Majority of customers are satisfied as they give a rating of 3 5.
- 3.1.3 Over 50% of customers who engaged with OCBC's 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:



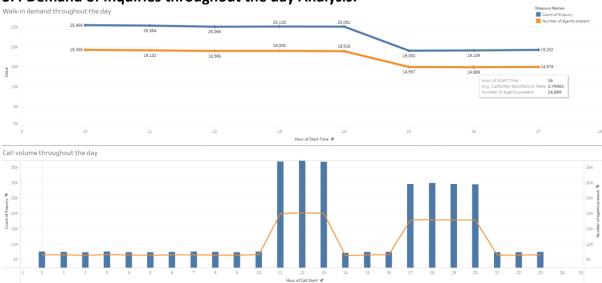
- 3.2.1 The bar chart displays the frequency of inquiry types, ordered by popularity. The **most common inquiries are related to bank transactions**, followed by services and card-related inquiries. This is true for both walk-in and call-in inquiries.
- 3.2.2 Breaking each inquiry down by satisfaction 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** for call-ins and walk-ins. The thinner bar for the call-in represents customers who did not leave feedback and does not represent a rating of 1.

# 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 a poor satisfaction rating average significantly extended waiting time. 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 has a poor average rating of 1-2. 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 an agent-to-customer 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.