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 :

**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). |
| [**https://chat.openai.com/c/3d949bfa-95f4-4601-b456-70b475e13c57**](https://chat.openai.com/c/3d949bfa-95f4-4601-b456-70b475e13c57) |
| 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> |

**Important Note:**

* Do not copy answers produced by the AI tool in totality as it is considered as plagiarism.
* Do not rely on any information produced by the AI tool blindly. You should always verify the answer with other sources. Do not assume that these answers provided by the AI tool are correct.
* To achieve quality outputs from the AI tool, you should provide good prompt that is clear and specific. Be precise and provide context. Avoid asking open-ended questions.

Content Page

Background and Business Requirements - 3

Data Dictionary -

**Task 1 : Background and Business Requirement**

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 **helpdesk** **staff 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 experiences** across different branches and customers.

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.

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.

The primary objectives of this analysis are as stated:

• What is the most common reason for calls or branch visits?

• Are there correlations between the feedback score and other factors?

• During which time periods do both channels experience their highest activity?

• Is there a specific time of day when one channel is preferred over the other?

**Task 2 : Data Management**

**Data Profiling**

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

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 |

Datasets Analysis

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**.

A screenshot of a computer screen

Description automatically generatedThere are 15 OCBC Branches, and the coordinates have no data quality issue.

Alteryx Designer was used for data inspection and cleaning. To discover potential quality issues for audit, I utilized the Field Summary Node to produce a report for each dataset.

[Screenshot of Branch Location Singapore on Excel]

Call-Log 2021:

A screenshot of a computer

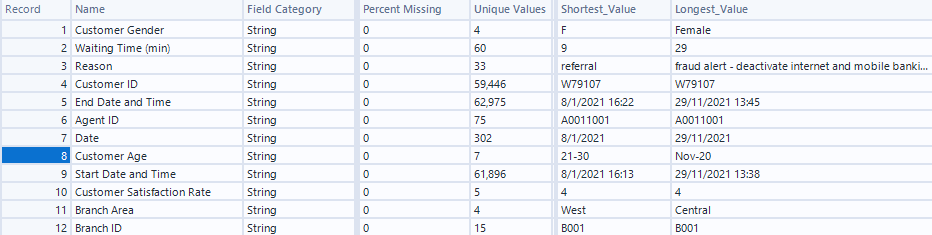
Description automatically generated

Call-Log 2022:

A screenshot of a computer

Description automatically generated

Walk-in 2021:



Walk-in 2022:

A screenshot of a computer

Description automatically generated

Data Quality Issues

A screenshot of a table

Description automatically generated

**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.

**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. A graph of blue bars

   Description automatically generated with medium confidenceGender label for customers in walk-in 2021 records is inconsistent as there are values **Female/Male when it should be F/M**.
4. 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. 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

   Description automatically generatedWalk-in 2022 has inconsistencies with Branch ID as there are Branch IDs that don’t start with B00.



Based on my analysis, I can confirm that the numbers represent the branch but it’s missing B00 as the branch area matches the branch area of its B000 + ID.



E.g. Branch ID 15’s branch area matches the branch area of Branch ID B015



**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 details, 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.

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

**Data Integration**

After performing the necessary data cleaning the address the data issues stipulated above 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 as 2022.

A screenshot of a computer

Description automatically generated A screenshot of a computer

Description automatically generated

Subsequently, I performed left join on the combined walk-in dataset to extract the longitude and latitude of the branch coordinates from the Branch Location Singapore dataset.

A screenshot of a computer

Description automatically generated

**Features Creation**

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 visualizations.

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

+ " " +

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

1. Duration

STR(DATEDIFF('second', [Call Start], [Call End]) / 60)

+ ":" +

STR(DATEDIFF('second', [Call Start], [Call End]) % 60)

**Task 3 : Data Exploration**

* Explain four (4) key takeaways based on your understanding of data. You will be assessed based on:
  + Correct usage of chart.
  + Demonstration of understanding of visualisation techniques.
  + Explanation of the purpose of each chart (consider composition, comparison, relationship, distribution).
  + Explanation of the relationships and analytical interaction.
  + How in-depth your analysis is and how sound your interpretation is.

|  |
| --- |
| SUBMISSION INSTRUCTIONS   * Save this file as “SD1\_[Your Class]\_[Your Full Name]\_[Your Student ID]”, e.g. “***SD1\_P01\_BATMAN ROBIN\_2299999A***”. * Submit your report in MS Word document to DAST LMS site > Assessment >   SD Part 1 - Data Visualisation Submission Link. |