Dashboard Link :
https://lookerstudio.google.com/reporting/edcc35cc-8fa0-4a58-9102-a6399fa7f7a1

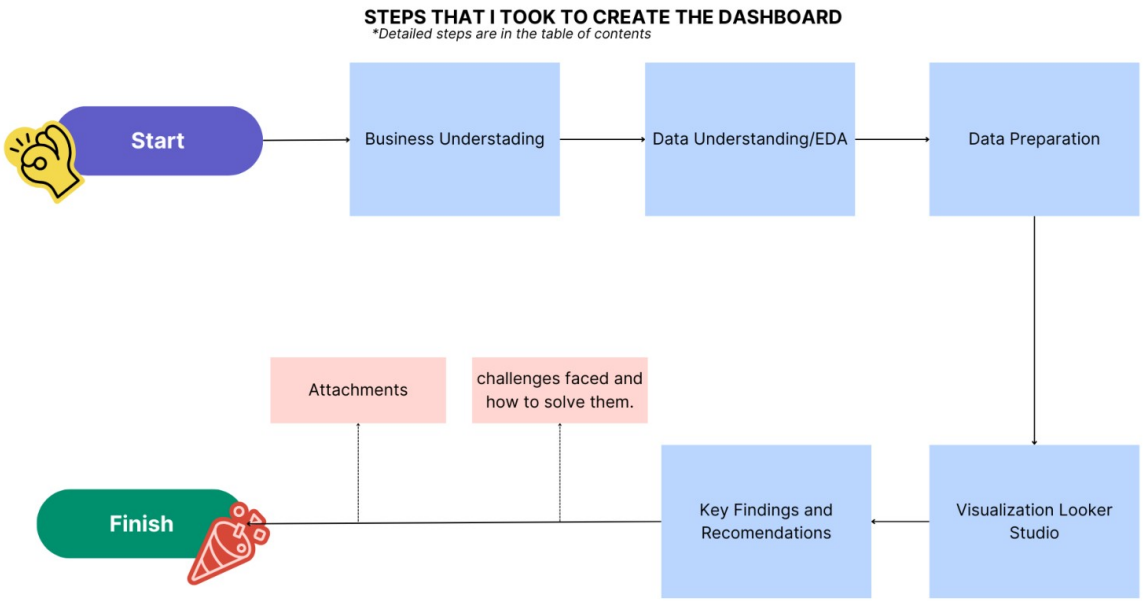
```
#import library yang dibutuhkan untuk mengerjakan tugas
import numpy as np
import pandas as pd
from PIL import Image
import matplotlib.pyplot as plt
```

---

## STEPS THAT I TOOK TO CREATE THE DASHBOARD

Detailed steps are in the table of contents

```
from PIL import Image
flow = Image.open(r'C:\Users\X1 CARBON\Downloads/Business Understading (6).png')
plt.figure(figsize=(10, 10), dpi=200)
plt.imshow(flow)
plt.axis('off')
plt.show()
```




# TABLE OF CONTENTS

5) Step 5 : Key Findings and Recomendations

6) Challenges faced and how to solve them.

7) Attachments

```
    - Field calculation (SQL Query) documentation from
field/column/metrics a
```

---

---

# STEP 1 : BUSINESS UNDERSTANDING

1. Identify Stakeholders needs ***(assumption)*** : The stakeholders need insights on user growth trends, survey participation, and payment trends to make informed decisions on platform strategy

2. Define Objective : Analyze user trends, survey participation, and payment trends to provide actionable insights that support strategic decision-making regarding platform growth, survey effectiveness, and finance.

3. Anaytical Approach : Descriptive(Primary), Diagnostic (Secondary)

1. Expected Business Metrics List :

**4.1 User Insights**

- Total Registered Users: The total number of users who have registered on the platform. This is a key metric for assessing the overall growth of the platform.

- New User Registration Trend: The rate at which new users are registering over time (e.g., per week or month). This helps to measure user acquisition and the effectiveness of marketing campaigns.

- User Demographic Distribution : The distribution of users across various demographic categories such as gender, education level, occupation, and city. This helps to understand the platform's user base composition and identify target groups for engagement.

**4.2 Survey Performance**

- Total Surveys Created : The total number of surveys created on the platform. This metric helps to measure the level of activity on the platform and the volume of surveys being launched.

- New Survey Creation Trend : The rate at which new surveys are created over time. This can indicate the growth and demand for surveys on the platform.

- Survey Completion Rate : The percentage of surveys that are completed by participants. This metric evaluates how successful surveys are in engaging participants and completing the survey process.
- Survey Enrollment Rate : The percentage of participants who enroll in a survey compared to the total number of participants needed. A higher enrollment rate indicates a well-targeted survey.
- Total Participants by Reward : Measures how participation rates vary depending on the reward offered for survey completion. This can reveal whether higher rewards correlate with higher participation rates.

### 4.3 Survey Participation

- New Participants Growth Trend : The rate at which new participants join surveys over time. This helps to track how well the platform is engaging users and attracting new participants.
- Total Participants : The total number of unique users who participate in surveys. This is an aggregate measure of user engagement with surveys across the platform.
- Demographic Participation Variations : The variation in survey participation rates across different demographic segments (e.g., gender, education level, occupation, city). This helps identify which groups are more engaged with surveys.

### 4.4 Payment Metrics

- Total Payment Volume : The total amount of money paid to users for participating in surveys. This is a financial metric that helps to track the cost of rewarding users and assess the platform's financial sustainability.
- Payment Volume Trend : The trend of payments made to users over time. This helps to track financial performance and determine whether payments are increasing as survey participation grows.
- Payment Volume by User Demographics : The distribution of total payment volume across different user demographics (e.g., gender, education, occupation). This helps to understand whether certain user groups are earning more or less in rewards, which can inform future targeting or reward strategies.

# STEP 2 : DATA UNDERSTANDING / EXPLORATORY DATA ANALYSIS

Objective: To find out the condition of the data and record the actionable points that must be carried out when processing data in Python or Looker using SQL Query so that the data is ready for visualization.

## 2.1 IMPORT DATA (COLLECT INITIAL DATA)

```
#Import payments.csv
payments = pd.read_csv(r'C:\Users\X1
CARBON\Downloads\Kudata/payments.csv')
#Import surveys.csv
surveys = pd.read_csv(r'C:\Users\X1
CARBON\Downloads\Kudata/surveys.csv')
#Import users.csv
users = pd.read_csv(r'C:\Users\X1 CARBON\Downloads\Kudata/users.csv')
#Import survey_participations.csv
participation = pd.read_csv(r'C:\Users\X1
CARBON\Downloads\Kudata/survey_participations.csv')
```

## 2.2 DESCRIPTIVE STATISTIC

Purpose: To view a statistical summary of the data.

1. Are there any missing values?
   - Method: Use **df.describe()**
   - Insight: There are no missing values in any of the dataframes.
2. Are there any column names that are incorrect (do not match the values/redundant)?
   - Method: Use **df.sample()**
   - Insight: The column **'created_at'** in the users dataframe and **'created_at'** in the surveys dataframe will cause a code error if both dataframes are joined. It is recommended to rename them to **'user_created_at'** and **'survey_created_at'**.
3. Are there any columns with incorrect data types?
   - Method: Use **df.info()**
   - Insight: The columns **'end_at', 'payment_date', 'response_date', and 'created_at'** should not be of type object. They should be datetime for easier data processing.

## 2.3 PREPARING DATA FOR METRICS VISUALIZATION (DATA REQUIREMENTS AND AVAILABILITY)

Purpose: Identify whether there is a need for a join/new column creation/new dataframe creation/other operations for a particular metric.

1. **User Insight**: No further metrics need to be processed; everything is fine.
2. **Survey Performance**:
   - **Survey Completion Rate**: Create a new field using the Field Calculation feature with an SQL query in Looker to calculate: (Total Number of Closed Surveys/Number of Surveys Reaching Participation Target)×100
   - **Survey Enrollment Rate**: Create a new field NAMED 'participation_rate' using the Field Calculation feature with an SQL query in Looker to calculate the Survey Participation Rate.

- **Total Participants by Reward**: Create a new field (reward_category) using the Field Calculation feature with an SQL query in Looker to categorize rewards into levels: low, medium, and high.

3. **Survey Participation**:
   - **Demographic Participation Variations**: Create a new dataframe called `merge_Part_Users` that contains the result of joining/merging the `participation` and `users` dataframes in Python (Alternatively, this can be done using the join table feature in Looker).
4. **Payments**:
   - **Payments Volume by User Demographics**: Create a new dataframe called `merge_Paym_Users` that contains the result of joining/merging the `payments` and `users` dataframes in Python (Alternatively, this can be done using the join table feature in Looker).

# STEP 3 : DATA PROCESSING

Objective: Process data so that it is ready to be visualized, including carrying out action points in step 2.|

## 3.1 Column Rename

1. Rename the column 'created_at' in the **users** table to 'user_created_at'.
2. Rename the column 'created_at' in the **surveys** table to 'surveys_created_at'.

```
#Rename kolom 'created_at' di pada users jadi 'user_created_at'
users.rename(columns={'created_at': 'user_created_at'}, inplace=True)
#Rename kolom 'created_at' di pada surveys jadi 'surveys_created_at'
surveys.rename(columns={'created_at': 'survey_created_at'},
inplace=True)
```

## 3.2 Change Data Types

1. Change the data types of 'end_at', 'payment_date', 'response_date', 'survey_created_at', and 'user_created_at' from object to datetime.

```
#ganti tipe data kolom 'end_at' dan 'survey_created_at' di dataframe
surveys
surveys['survey_created_at'] =
pd.to_datetime(surveys['survey_created_at'])
surveys['end_at'] = pd.to_datetime(surveys['end_at'])

#ganti tipe data kolom 'payment_date' di dataframe payments
payments['payment_date'] = pd.to_datetime(payments['payment_date'])

#ganti tipe data kolom 'response_date' dan 'survey_created_at' di
dataframe pasticipations
participation['response_date'] =
pd.to_datetime(participation['response_date'])
```

```
#ganti tipe data kolom 'response_date' dan 'survey_created_at' di
dataframe pasticipations
users['user_created_at'] = pd.to_datetime(users['user_created_at'])
```

3.3 Creating New DataFrames

1. **Metrics Demographic Participation Variations**: Create a new dataframe named `merge_Part_Users` that contains the result of joining/merging the `participation` and `users` dataframes in Python.

2. **Metrics Payments Volume by User Demographic**: Create a new dataframe named `merge_Paym_Users` that contains the result of joining/merging the `payments` and `users` dataframes in Python.

```
# Membuat merge_part_users sebagai pengganti dataframe participation
untuk kedepannya
merge_part_users = pd.merge(participation, users, on='user_id',
how='left')
merge_part_users = merge_part_users.drop(['full_name', 'email',
'user_created_at'], axis=1)
merge_part_users =
merge_part_users.sort_values(by='response_date',ascending=True)

# Membuat merge_paym_users sebagai pengganti dataframe payments untuk
kedepannya
merge_paym_users = pd.merge(payments, users, on='user_id', how='left')
merge_paym_users = merge_paym_users.drop(['full_name', 'email',
'user_created_at'], axis=1)
merge_paym_users =
merge_paym_users.sort_values(by='payment_date',ascending=True)
```

3.4 The dataframes `merge_part_users`, `merge_paym_users`, and those that have been processed should be exported to CSV for visualization in Looker Studio.

```
# Export merge_part_users ke file csv
merge_part_users.to_csv(r'C:\Users\X1
CARBON\Downloads\Kudata/merge_part_users.csv', index=False)

# Export merge_part_users ke file csv
merge_paym_users.to_csv(r'C:\Users\X1
CARBON\Downloads\Kudata/merge_paym_users.csv', index=False)

# Export user yang sudah di feature transformation ke file csv
users.to_csv(r'C:\Users\X1 CARBON\Downloads\Kudata/users_ft.csv',
index=False)

# Export surveys yang sudah di feature transformation ke file csv
surveys.to_csv(r'C:\Users\X1 CARBON\Downloads\Kudata/surveys_ft.csv',
index=False)
```

# STEP 4 : Visualization Using Looker Studio

Link : https://lookerstudio.google.com/reporting/edcc35cc-8fa0-4a58-9102-a6399fa7f7a1

# STEP 5 : KEY FINDINGS AND RECOMENDATION

## Key Findings

**1. User Insights**

- **Total Registered Users**: Currently, there are **1,000 registered users** on the platform.
- **New User Registration Trend**: New user growth shows **stagnant fluctuations**, with an average increase of **2.62 users** per month over the period from 2003 to 2024. This indicates that there has been **no significant progress in acquiring new users**, which could be an area of concern for marketing and user acquisition strategies.
- **User Demographics**: The distribution of users based on **city, occupation, gender, and education** is relatively **even**. This indicates that there is **no dominant or unbalanced demographic category**, meaning the platform is reaching various user groups effectively.

---

**2. Survey Performance**

- **Total Surveys**: There are **200 surveys** that have been created on the platform. However, the **growth of surveys** from 2023 to 2024 **appears to be stagnant**, in line with user growth. This suggests that although there is an increase in users, the **activity in survey creation** needs to be improved to keep the platform dynamic.
- **Survey Completion Rate**: No surveys have successfully **met the participation target**. On average, surveys only managed to gather **48.4%** of the required participants. This is a **significant indicator of issues with user engagement**, which could affect the quality of surveys and the results obtained.
- **Survey Enrollment vs. Reward Correlation**: There is a **significant positive correlation** between the **size of the reward** and the **survey participation rate**. Data shows that the higher the reward offered, the more participants are interested in joining. This suggests that **financial incentives** play a crucial role in attracting user participation in surveys.

---

**3. Participation Analysis**

- **Total Participants**: **5,000 users** have participated in surveys.
- **New Participants Growth Trend**: From 2023 to 2024, there has been a **fluctuating increase in the number of new participants** filling out surveys daily. This indicates that there is potential to **increase participation more consistently**, although fluctuations need further analysis.
- **Participant Demographics**: Like the **user demographics**, the distribution of participants based on **city, occupation, gender, and education** is also **even**, with no dominance in any single category. This indicates that survey participation is well-distributed among various user segments, although further research is needed to understand why there are fluctuations in daily participation numbers.

**4. Payment Insights**

- **Total Payment Volume**: A total of **131.3K** has been paid to survey participants during the 2023-2024 period. This reflects a **substantial payment volume**, but further analysis is needed regarding the **daily payment fluctuations**, which appear to decrease and lack consistency.
- **Payment Trends**: Daily payments show **significant fluctuations**, which could indicate irregularities in payment flows or an imbalance in payment distribution based on survey participation or performance.
- **Payment Volume by Demographics**: Similar to participant and user distributions, payments to participants are also **evenly distributed** based on **city, occupation, gender, and education**, indicating that the payment system is **fair**, with no group being disproportionately favored or disadvantaged in terms of compensation.

### Recommendations

1. Focus on creating surveys with high rewards, and reduce surveys with low rewards to save costs, as there is an increase in participation with higher rewards.
2. Since there is a correlation between the number of users creating accounts and the number of surveys created, to attract new users, it is necessary to create more and more relevant surveys.
3. Allocate the budget towards surveys with high rewards to encourage more participation.
4. Reduce surveys with low rewards to save costs and maintain focus on more effective surveys.
5. Optimize survey design to make it more engaging and easier to follow, in order to improve the completion rate.
6. Increase targeted marketing campaigns based on user demographics to attract more participants from specific groups. Focus on certain niches to make marketing costs more efficient.
7. Align payments with performance to ensure that payments are more efficient and fair based on user participation. ansial.

## 6. Challenges faced and how to solve them.

**1. Identifying Representative Metrics**

- **Challenge**: It was difficult to select metrics that directly aligned with stakeholders' needs, focusing on user growth, survey participation, and payment trends.
- **Solution**: I identified key metrics such as **Total Registered Users, Survey Completion Rates**, and **Payment Volumes** that align with business objectives and provide actionable insights for decision-making.

**2. Data Integration and EDA**

- **Challenge**: Performing **EDA** to determine which data needed to be joined and processed was complex, given the multiple tables (Users, Surveys, Participation, Payments).

- **Solution**: I conducted thorough EDA to assess data structure, cleaned the data, and merged tables on common identifiers (e.g., `user_id`, `survey_id`) to ensure consistency and completeness.

**3. Finding the Right Visualizations**

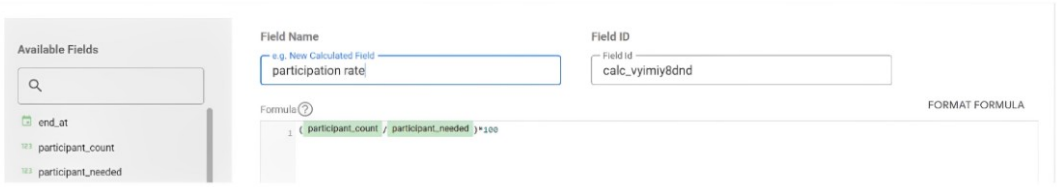
- **Challenge**: Selecting the right visualizations to represent diverse metrics in a clear and actionable way was challenging.
- **Solution**: I used **line charts** for trends, **bar graphs** for comparisons, and **heatmaps** for demographic insights, ensuring interactivity for stakeholders to explore specific segments.

# 7 : ATTACHMENTS

```
# Dokumentasi Field Calculation SQL Query di Looker Studio
import matplotlib.pyplot as plt
import matplotlib.image as mpimg

image_path = r'C:\Users\X1 CARBON\Downloads\attachment.png'

img = mpimg.imread(image_path)

fig = plt.figure(figsize=(10, 8), dpi=200)
plt.imshow(img)
plt.axis('off')
plt.show()
```

1. Reward_Category Column for Total Participants by Reward Metrics


2. Participation_rate field for Survey Enrollment Rate Metrics




3. completion_rate field for survey completion rate Metrics