

# InsideOut

## Data Analysis Final Project – DEPI

Prepared by:

1-Hager Lialy

2-Sandy Nazeh

3-Dina Mohamed

4-Omnya Talaat

## 1. Abstract

The project aims to analyze various aspects of InsideOut, a small business, to enhance sales and improve customer satisfaction. The focus is on understanding customer demographics, behavior, and satisfaction to assess brand positioning and engagement. Additionally, the study evaluates the effectiveness of marketing campaigns across three platforms (Facebook, Instagram, and TikTok) by analyzing content types, post-performance metrics (likes, comments, shares, views), and their impact on sales and engagement. The analysis also explores sales performance by examining order frequency, product preferences, and the influence of dollar exchange rates on sales.

A combination of a structured SQLite database, Python, Excel, and Tableau, along with survey data, is used to generate insights. The results will provide actionable recommendations to optimize marketing strategies, improve customer satisfaction, and drive sales growth.

## 2. Introduction

### Project Overview

In today's competitive market, businesses must continuously adapt their strategies to maintain customer engagement, optimize marketing campaigns, and increase sales. This project is an integrated effort to combine sales and marketing data with customer survey responses to offer a comprehensive evaluation of business performance across multiple fronts. The project is split into two primary aspects: an internal database that houses marketing and sales data, and a customer survey designed to gauge customer behavior, brand perception, and satisfaction.

### Objectives

The main objectives of this project are to:

- Analyze Customer Demographics, Behavior & Satisfaction
- Analyze Brand Edge: Measure the competitive edge of the brand through customer feedback on brand loyalty and why do they choose it.
- **Evaluate Sales Performance:** Examine sales trends, product preferences, order frequency by gender and location.
- **Assess Dollar Change Rate Impact:** Analyze the relationship between dollar price fluctuations and sales, determining how changes in currency value affect purchasing decisions and overall revenue.
- **Impact of Marketing Campaigns:** Assess the effectiveness of marketing campaigns across Facebook, Instagram, and TikTok. The analysis includes metrics such as likes, comments, shares, and views, with filters applied to content type, post type, and platform to determine which combinations yield the highest engagement.
- **Data-Driven Strategic Recommendations:** Provide actionable insights based on data analysis to optimize future marketing strategies, enhance customer engagement, and improve sales performance.

## 3. Methodology

### 3.1. Data sources & collection

The data for this analysis was gathered from two primary sources:

### 3.1.1 Internal Database (Marketing & Sales Data):

- Sales data was initially extracted from InsideOut's Excel sheets, subsequently cleaned, organized, and converted into CSV files before being imported into the SQLite database.
- The marketing data was collected manually from social media platforms insights.
- Dollar Exchange data collected from historical data from the Central Bank of Egypt.

3.1.2 Survey Data: Survey data was collected directly from customers through individualized surveys, aimed at assessing customer behavior, brand perception, and overall satisfaction.

## **3.2. Data Description**

The dataset consists of 10 tables, each playing a vital role in the analysis:

### **1. Product Table:**

Contains information about the products offered by InsideOut.

- **Columns:** Product\_ID, Product\_Type
- **Number of rows:** 736

### **2. Customers Table:**

Stores customer details.

- **Columns:** Customer\_ID, Gender, Location
- **Number of rows:** 736

### **3. Orders Table:**

Tracks sales data.

- **Columns:** Date\_ID, Order\_ID, Customer\_ID, Product\_ID, Order\_Status, No\_of\_products, Product\_cost, Product\_price, Total\_Price, Dollar\_exchange\_rate
- **Number of rows:** 736

### **4. Marketing Table:**

Contains marketing campaign data from Facebook, Instagram, and TikTok, tracking various metrics.

- **Columns:** Date\_ID, Post\_type, Content\_type, Likes, Shares, Comments, Views, Platform\_ID
- **Number of rows:**
  - Instagram: 181

- TikTok: 117
- Facebook: 153

### 5. Platforms Table:

Stores information about social media platforms.

- **Columns:** Platform\_ID, Platform\_name
- **Number of rows:** 3

### 6. Date Table:

Stores date-related data.

- **Columns:** Date\_ID, Date
- **Number of rows:** 27

### 7. Gender Table:

Contains follower demographics on Instagram and Facebook.

- **Columns:** Platform\_ID, Gender, Percentage
- **Number of rows:** 5

### 8. Age Table:

Contains follower age demographics on Instagram and Facebook dashboards.

- **Columns:** Platform\_ID, Age, Percentage
- **Number of rows:** 13

### 9. Countries Table:

Stores country-based follower demographics on Instagram and Facebook dashboards.

- **Columns:** Platform\_ID, Country, Percentage
- **Number of rows:** 11

### 10. Cities Table:

Contains city-based follower demographics on Instagram and Facebook dashboards.

- **Columns:** Platform\_ID, City, Percentage
- **Number of rows:** 5

### 11. Customer Feedback Table:

Contains survey data capturing customer feedback.

- **Columns:**

- Age\_Group: Represents age categories (Under 18, 18-24, 25-34, 35 or older).
- Gender: Binary data (Male, Female).
- Location: Represents customer locations (Cairo, Giza, Alexandria, الدلتا, صعيد مصر).
- Occupation: Represents customer occupations (Student, Employee, Both, None).
- Platform\_Preference: Indicates which platform customers prefer for the brand (Instagram, Facebook, TikTok).
- First\_Source\_of\_Hearing: How customers first heard about the brand (Instagram, Facebook, TikTok, Friend, Influencer, Brand Owner).
- Most\_Valued\_Feature: What customers value most in the planner (Design, Size, Structure, Other).
- Willingness\_to\_Pay: Customer's willingness to pay for a high-quality planner (200-250 EGP, 250-350 EGP, 350-450 EGP, 450-550 EGP).
- Interest\_in\_Additional\_Products: Multi-select data indicating interest in other products (Notebooks, Planner Accessories, Memos, To-do Lists, Weekly Planners, Not Interested).
- Willingness\_to\_Buy\_as\_Gift: Whether customers would buy the planner as a gift (Yes, No, Maybe).
- Preferred\_Promotions: Type of promotion that would encourage a purchase (Discount Codes, Buy One Get 30% Off, Free Shipping, Limited Edition Designs, Loyalty Programs).
- Reason\_for\_Not\_Purchasing: Reasons for not purchasing yet (Shipping Costs, Unappealing Designs, Product Price, Not Sure How to Organize Time, Other).
- Preferred\_Time\_to\_Purchase: Ideal time to buy a planner (Start of a New Year, Beginning of School Semester, No Specific Time).
- Has\_Purchased\_Planner: Indicates whether the customer has purchased a planner (Yes, No).
- Issues\_with\_Planner: Issues encountered with the planner (Quality, Shipping Problems, Incorrect Order, Other).
- Satisfaction\_with\_Purchase: How satisfied customers are with their purchase (Very Satisfied, Satisfied, Not Satisfied, None).
- Likelihood\_of\_Purchasing\_Again: Indicates the likelihood of making another purchase (Yes, No, None).

- **Number of rows:** 245

### 3.3. Tools

The project was conducted from A to Z on 5 different tools, but each tool had different specific tasks including:

- **SQL & SQLite:** Employed to design and query the database schema, enabling the extraction of insights and generation of reports based on sales and marketing data.
- **Python (Pandas, Matplotlib, NumPy, Seaborn):** used for data cleaning, in-depth analysis, and visualization.
- **Excel:** applied for manual data management, summary statistics, and additional data visualizations.
- **Tableau:** used to convert raw data into visual insights through interactive dashboards, comprehensive reports, and charts.
- **R:** Used for data analysis and visualization, including:
  - **dplyr:** For data manipulation.
  - **ggplot2:** For creating visualizations.
  - **DBI:** For database management.
  - **RSQLite:** For handling SQLite databases.
  - **readr:** For reading and writing data.
  - **tidyr:** For tidying data.
  - **patchwork:** For combining plots.
  - **plotly:** For interactive graphs.

### 3.4. Methods

#### 3.4.1. Data Preprocessing:

Before conducting the analysis, the data from the internal database and customer surveys underwent several preprocessing steps:

- **Data Cleaning:** The raw data was checked for missing values, duplicates, and inconsistencies by using excel, python, and . Records with significant missing data were either imputed using appropriate methods or removed if they could not be accurately reconstructed.
- **Normalization:** Numerical variables were normalized to handle outliers. Median values were used for analysis to minimize the effect of extreme values.
- **Categorization:** Survey responses were categorized to group similar responses together, which allowed for more robust analysis of customer feedback.

**3.4.2. Database Construction:** We built the schema in SQLite, connecting the 9 tables based on ID fields.

### 3.4.3 Exploratory Data Analysis Techniques:

#### Marketing Campaign Analysis:

- **Metrics:** Relationships between content type, post type, and engagement metrics (likes, comments, shares, views) were analyzed across three platforms: Facebook, Instagram, and TikTok. Filters were applied to segment results by platform and content type.

#### Sales Analysis:

- **Order Analysis:** The number of orders for different product types, as well as the total sales for specific periods (two months), were calculated. Sales performance was broken down by gender, location, and product type to gain deeper insights into purchasing behavior.
- **Dollar Exchange Rate Analysis:** The relationship between the dollar exchange rate and sales performance was analyzed over multiple years to determine the impact of currency fluctuations on product costs and customer purchasing behavior.

#### Demographic Analysis:

- **Customer Distribution:** Demographic data from both the internal database and the customer survey was visualized to provide insights into the geographic distribution of customers, their age groups, and gender breakdown across the platforms.

#### Impact of Marketing on Sales:

- The effect of marketing campaigns on sales performance was analyzed by examining the correlation between the total number of views (across all platforms) and the number of products sold per month. Additional analysis looked at the frequency of different selling



content posted each month and the number of products sold, identifying which content types had the most influence on sales.

#### Survey Data Analysis:

- The survey data was analyzed to identify patterns in customer satisfaction and engagement. Insights from the survey were combined with the sales and marketing data to draw conclusions about the effectiveness of marketing strategies and the overall brand perception.

#### 3.4.5: Visualization:

Final dashboards were created in Excel & Tableau showing insights summary

## 4. Contribution

This project contributes significantly to Inside Out's marketing and sales strategies by providing actionable insights. Key contributions include:

#### Sales Insights

##### 1. Gender Distribution of Sales

- **Insight:** Female customers account for the overwhelming majority of sales, while male engagement is very low.
- **Recommendation:** Implement a new marketing strategy tailored to male customers. This could involve promoting planner features that resonate with male audiences or creating campaigns with male influencers.

##### 2. Product Performance (1-Year vs. 6-Month Planners)

- **Insight:** The 1-Year Planner vastly outperforms the 6-Month Planner in terms of sales.
- **Recommendation:** Focus resources on promoting the 1-Year Planner, as it's clearly preferred. For the 6-Month Planner, consider repositioning or rebranding it to appeal to a niche market, such as professionals or students with shorter-term planning needs.

##### 3. Cancellation Analysis

- **Insight:** There were significant cancellations in December 2023 due to major delays in order fulfillment, particularly during the holiday rush. Other unidentified reasons may also have contributed.
- **Recommendation:** Strengthen the logistics and order fulfillment process during high-demand periods. Improve communication with customers regarding

shipping delays and consider offering incentives (e.g., discounts on future purchases) to reduce cancellations.

#### 4. Geographic Distribution of Sales

- **Top Regions:** Cairo and Giza dominate sales, indicating strong market penetration. Alexandria is also performing reasonably well.
- **Low Performing Areas:** Locations like Wady Elgdad, Sharm Elsheikh, and Mansoura show minimal sales.
- **Recommendation:** While Cairo and Giza can maintain standard marketing strategies, low-performing areas need targeted marketing efforts, possibly through localized promotions or collaborations with regional influencers to boost sales.

#### 5. Seasonal Trends

- **Insight:** December consistently sees the highest order volume, especially during the holiday season.
- **Recommendation:** Increase marketing efforts leading up to December to capitalize on seasonal shopping trends.

### Marketing Insights

#### 1. Content Engagement

- **Insight:** Engagement varies by content type, with storytelling and selling content generating the most interaction.
- **Recommendation:** Prioritize storytelling and selling content in marketing campaigns to maximize audience engagement.

#### 2. Content Formats

- **Insight:** Reels outperform other formats in engagement.
- **Recommendation:** Focus on creating more reels to leverage their popularity and enhance customer interaction.

#### 3. relationship between selling content posts and orders

- **Insight:** There is a direct relationship between the number of selling content posts and orders.

- **Recommendation:** Implement promotional campaigns during slower months to boost sales.

#### 4. Views vs. Orders

- **Insight:** Higher view counts generally lead to more orders, but some months have high views without corresponding sales.
- **Recommendation:** Improve conversion strategies, such as clear calls-to-action, to turn views into sales during peak view months.

#### 5. Audience Engagement

- **Insight:** Female customers engage more than male customers.
- **Recommendation:** Develop targeted campaigns to attract male customers and diversify the audience base.

### Marketing Insights on Anomaly Detection

#### 1. Facebook Anomalies

The following posts have been identified as anomalies based on views:

- **Content Types:** Selling content, Storytelling, and Awareness content.
- **Observation:** The anomalies indicate posts that received significantly higher engagement than the average.
- **Recommendation:** The business owner should analyze these high-performing posts to understand what strategies contributed to their success. This can guide future content creation and marketing efforts, particularly focusing on elements that resonate with the audience.

#### 2. Instagram Anomalies

The identified anomalies on Instagram show substantial views:

- **Content Types:** Selling content, Storytelling, and Awareness content.
- **Observation:** Posts with exceptionally high views could signal effective messaging or promotion strategies.
- **Recommendation:** Investigate the common characteristics of these anomaly posts, including content format, messaging, and posting time. Leverage successful elements in upcoming campaigns to enhance overall engagement.

#### 3. TikTok Anomalies

The TikTok platform has its share of notable anomalies in views:

- **Content Types:** Selling content and Storytelling.
- **Observation:** Posts that stand out indicate potential viral content or trending topics that resonated with viewers.
- **Recommendation:** The business owner should consider increasing focus on the types of content that led to these anomalies. Identifying trends or themes could provide valuable insights for future TikTok campaigns, optimizing engagement and reach.

## Customer Feedback Insights

### Customer Overview

#### 1. Demographics

- **Location Distribution:** Majority of customers are from Cairo, followed by Delta, Giza, and Alexandria.
- **Recommendation:** Tailor marketing campaigns to focus on these regions, particularly Cairo, where customer engagement is highest.

#### 2. Willingness to Pay

- Most customers willing to pay between 250-350 EGP.
- **Recommendation:** Consider pricing strategies within this range and explore value-added options that justify higher prices, such as premium features or bundled offers.

#### 3. Platform Suitability

- **Preferred Platforms:** Instagram is most suitable, followed by TikTok and Facebook.
- **Recommendation:** Prioritize marketing efforts on Instagram and TikTok, utilizing influencers to enhance brand visibility and engagement.

#### 4. Promotion Preferences

- Customers respond well to discount codes and free shipping.
- **Recommendation:** Implement regular promotional campaigns featuring discount codes and free shipping offers, especially during peak purchasing seasons.

#### 5. First Awareness

- Influencers and Instagram are crucial for brand awareness.

- **Recommendation:** Collaborate with relevant influencers to enhance brand reach and create targeted ads on Instagram.

## Brand Edge

### 1. Product Interests

- Strong interest in to-do lists and weekly planners.
- **Recommendation:** Focus product development and marketing on these items, possibly offering customization options.

### 2. Timing of Purchases

- Peak purchases at the start of the new year.
- **Recommendation:** Launch targeted campaigns during the New Year season, promoting planners as essential tools for new beginnings.

### 3. Purchase Behavior

- Significant number of customers have not purchased planners.
- **Recommendation:** Implement strategies to convert hesitant customers, such as limited-time offers or testimonials highlighting customer satisfaction.

### 4. Favorite Features

- Inner structure of planners is highly valued.
- **Recommendation:** Highlight the unique features of planners in marketing materials and provide samples or demonstrations to showcase functionality.

### 5. Purchase Barriers

- Price is a primary barrier, especially for students.
- **Recommendation:** Introduce student discounts or loyalty programs that reward frequent purchases.

## Customer Satisfaction

### 1. Shipping Concerns

- Delta customers express dissatisfaction with shipping costs.
- **Recommendation:** Review and optimize shipping strategies for this region, potentially exploring partnerships with local couriers for better rates.

### 2. Problem Encounters

- Quality issues reported frequently.
- **Recommendation:** Strengthen quality control processes and consider gathering customer feedback post-purchase to identify recurring issues.

### 3. Satisfaction Levels

- Majority are very satisfied.
- **Recommendation:** Continue to engage satisfied customers through loyalty programs or referral discounts to encourage word-of-mouth marketing.

### 4. Repurchase Intent

- Strong intent to repurchase exists.
- **Recommendation:** Develop follow-up marketing strategies to remind customers to repurchase, possibly offering exclusive deals for repeat buyers.

### 5. Satisfaction Despite Issues

- Some customers remain satisfied despite problems.
- **Recommendation:** Investigate how to turn these customers into advocates; consider personalized follow-ups or incentives for providing testimonials.

### 6. Price Sensitivity

- Students most affected by price barriers.
- **Recommendation:** Create targeted marketing campaigns specifically addressing students, including budget-friendly options or flexible payment plans.

## 5. Team Members' Contributions

This section outlines the contributions of each team member throughout the project:

### Data Collection

- All team members

### Survey

- **SQL:** Sandy Nazeh
- **Python:** Dina Mohamed
- **Excel:** Omnia Talaat, Sandy Nazeh
- **Excel Dashboard:** Hager Lialy

- **Tableau:** Hager Lialy

### **Marketing & Sales**

- **SQL:** Hager Lialy, Sandy Nazeh
- **Python:** Hager Lialy, Sandy Nazeh
- **Excel:** Dina Mohamed
- **Excel Dashboard:** Sandy Nazeh, Dina Mohamed
- **Tableau:** Dina Mohamed
- **R:** Omnia Talaat

### **Documentation**

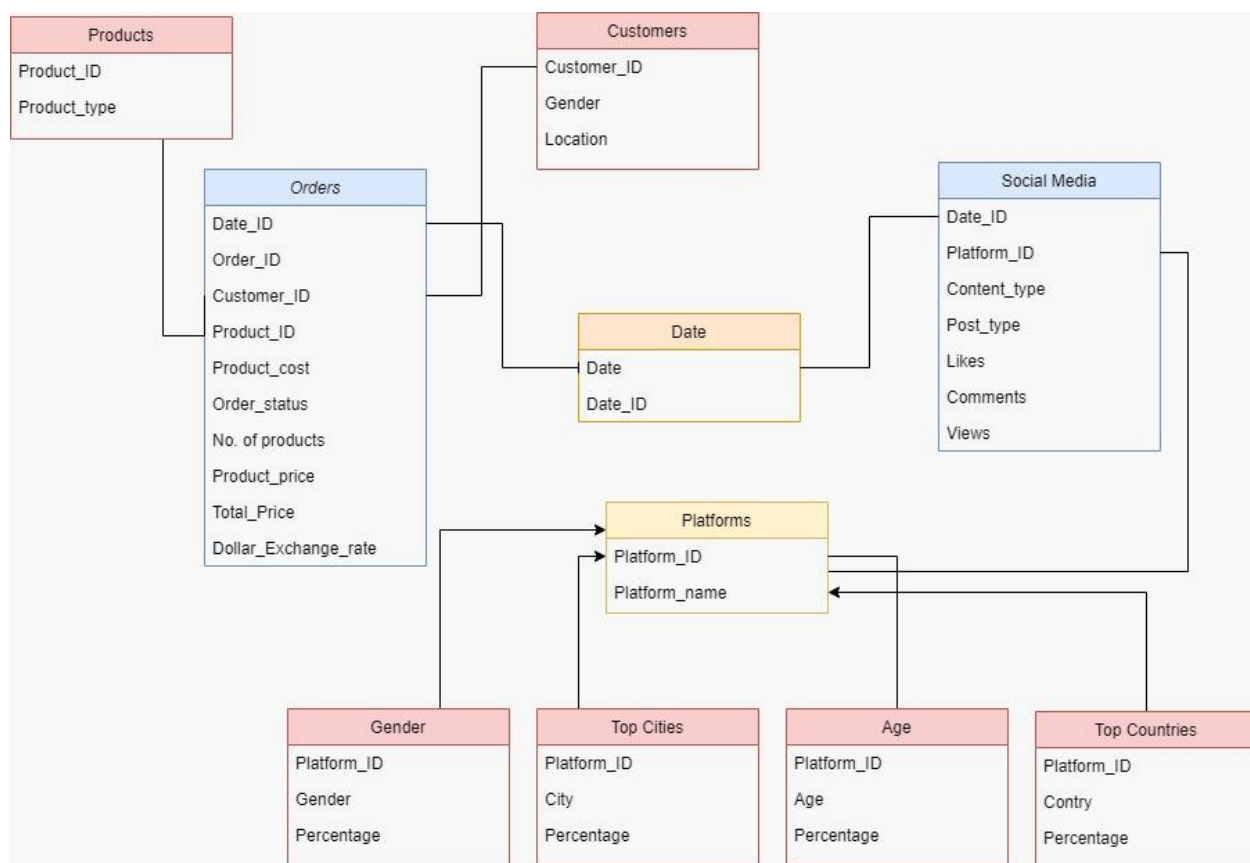
- Dina Mohamed, Hager Lialy, Omnia Talaat, Sandy Nazeh

### **Presentation**

- Dina Mohamed

## 6. Appendix

### 6.1. project schema



### 6.2. Data Collection

#### 6.2.1 Manual collection for marketing data for each platform



Date	Content type	Post type	No. of likes	No. of comments	No. of shares	No. of views
Sep-24	selling content	Post	10	5	1	186
Aug-24	Storytelling	Reel	9	0	0	245
Aug-24	Storytelling	Reel	8	0	0	264
Aug-24	Storytelling	Reel	6	1	0	223
Jun-24	Storytelling	Reel	18	6	0	663
Jun-24	educational content	Reel	7	2	0	392
Jun-24	selling content	Carousel post	4	0	0	184
Jun-24	Influencers	Reel	9	2	0	653
Jun-24	selling content	Post	3	0	0	160
May-24	selling content	Post	3	0	0	125
Mav-24	selling content	Carousel post	12	4	0	203

6.2.2. Final marketing sheet after adding Date\_ID & Platform\_ID for merging all 3 platforms together and to link them to database.

Date_ID	Content_type	Post_type	likes	comments	shares	views	Platform_ID
92024	selling content	Post	10	5	1	186	111
82024	Storytelling	Reel	9	0	0	245	111
82024	Storytelling	Reel	8	0	0	264	111
82024	Storytelling	Reel	6	1	0	223	111
62024	Storytelling	Reel	18	6	0	663	111
62024	educational cont	Reel	7	2	0	392	111
62024	selling content	Carousel pos	4	0	0	184	111
62024	Influencers	Reel	9	2	0	653	111
62024	selling content	Post	3	0	0	160	111
52024	selling content	Post	3	0	0	125	111
52024	selling content	Carousel pos	12	4	0	203	111

6.2.3. Initial sales sheet from the business owner

Order ID	Column1	Column2	Name	Platform	Account Link	Column3	Column4	Type	Item
161.1	Female	Delivered	Amira Jaheen	Instagram	<a href="https://www.instagram.com/amirajahen?utm_source=ig_web_profile_page">https://www.instagram.com/amirajahen?utm_source=ig_web_profile_page</a>			Book + Pastel	2
162	Female	Delivered	Khadija Ahmed	Instagram	<a href="https://www.instagram.com/khadija_ahmed?utm_source=ig_web_profile_page">https://www.instagram.com/khadija_ahmed?utm_source=ig_web_profile_page</a>			White + Quarters	1
163	Female	Delivered	Malaz	Instagram	<a href="https://www.instagram.com/malaz_ahmed?utm_source=ig_web_profile_page">https://www.instagram.com/malaz_ahmed?utm_source=ig_web_profile_page</a>			Book + Pastel	1
164	Female	Cancelled	Nada Rashad	Instagram	<a href="https://www.instagram.com/nadashad?utm_source=ig_web_profile_page">https://www.instagram.com/nadashad?utm_source=ig_web_profile_page</a>				

6.2.4 The sales sheet was divided into 3 sheets (Products, Customers and Orders) to link them together and with marketing sheet in database.

- Products sheet:

Product_ID	Product_Type
1	Planner (1 year)
2	Planner (6 months)

- Orders sheet:

Date_ID	Order_ID	Customer_ID	Product_ID	Order_Status	No_of_products	Product_cost	Product_price	Total_Price	Dollar_exchange_rate
122022	101	1	1	Delivered	1	85	130	130	24.6
122022	201	2	1	Delivered	1	85	130	130	24.6
122022	301	3	1	Delivered	1	85	130	130	24.6
122022	401	4	1	Delivered	1	85	130	130	24.6
122022	501	5	1	Delivered	1	85	130	130	24.6
122022	601	6	1	Delivered	1	85	130	130	24.6
122022	701	7	1	Delivered	1	85	130	130	24.6
122022	801	8	1	Delivered	1	85	130	130	24.6
122022	901	9	1	Delivered	1	85	130	130	24.6
122022	1001	10	1	Delivered	1	85	130	130	24.6
122022	1101	11	1	Delivered	1	85	130	130	24.6
122022	1201	12	1	Delivered	1	85	130	130	24.6

-Customers sheet:

Customer_ID	Gender	Location
1	Female	Cairo
2	Female	Cairo
3	Female	Cairo
4	Female	Cairo
5	Male	Cairo
6	Female	Cairo
7	Female	Cairo

6.2.5: The extracted sheet of survey's data from google form:

Timestamp	What is your age group?	What is your gender?	Where are you located?	What is your current occupation?	Which platform do you believe is most suitable?	How did you first hear about InsideOut?	What is the feature?
9/11/2024 19:37:36	Under 18	Female	Cairo	Student	Instagram	Instagram	Design & Style
9/11/2024 19:40:06	18-24	Female		Student	Instagram	Instagram	Design & Style
9/11/2024 19:40:52	18-24	Female	Cairo	Both	Instagram	Friend	Inner Structure (All)
9/11/2024 19:41:01	18-24	Female	Cairo	Student	Instagram	Influencer	Inner Structure (All)
9/11/2024 19:41:26	18-24	Female	Cairo	Student	Instagram	Instagram	Inner Structure (All)
9/11/2024 19:42:52	18-24	Female	Cairo	Student	Instagram	Friend	Inner Structure (All)
9/11/2024 19:44:06	18-24	Female	Giza	Student	Instagram	Instagram	Inner Structure (All)
9/11/2024 19:46:53	Under 18	Female	Cairo	Student	Instagram	Instagram	Inner Structure (All)
9/11/2024 19:47:45	18-24	Female	Cairo	Student	Instagram	Instagram	Inner Structure (All)

6.2.6: The cleaned survey's data sheet:

Location	Occupation	Platform Suitability	First Heard	Planner Feature	Willing to Pay	Interested Products	Purchase Gift	Promotion Encouragement	Purchase
Cairo	Student	Instagram	Instagram	Design & Style	200-250 EGP	Weekly planners	Yes	Free shipping	I'm not
Delta	Student	Instagram	Instagram	Design & Style	250-350 EGP	Notebooks, Planner accessories (e.g., stickers, pens), Memos, To-do lists	Yes	Free shipping	I'm not
Cairo	Both	Instagram	Friend	Inner Structure (All)	250-350 EGP	Notebooks, Planner accessories (e.g., stickers, pens), Memos, To-do lists, Weekly planners	Yes	Loyalty programs	null
Giza	Student	Instagram	Instagram	Inner Structure (All)	200-250 EGP	To-do lists	Maybe	Discount codes	Product
Cairo	Student	Instagram	Instagram	Inner Structure (All)	200-250 EGP	Notebooks, Planner accessories (e.g., stickers, pens), Weekly planners	Yes	Discount codes	Product
Cairo	Employee	Instagram	Instagram	Design & Style	200-250 EGP	Planner accessories (e.g., stickers, pens), Weekly planners	Maybe	Discount codes	Product
Cairo	Student	Instagram	Friend	Inner Structure (All)	250-350 EGP	To-do lists, Weekly planners	Yes	Discount codes	null
Cairo	Student	Instagram	Instagram	Size & Ease of Use	250-350 EGP	Notebooks, Planner accessories (e.g., stickers, pens)	Yes	Discount codes	Unappealing
Delta	Student	Instagram	Friend	Inner Structure (All)	250-350 EGP	Notebooks, To-do lists	Yes	Free shipping	I'm not
Giza	Both	TikTok	Friend	Inner Structure (All)	250-350 EGP	Planner accessories (e.g., stickers, pens), To-do lists	Yes	Free shipping	null
Cairo	Student	Instagram	Instagram	Inner Structure (All)	350-450 EGP	Notebooks, Planner accessories (e.g., stickers, pens), Memos, To-do lists, Weekly planners	Maybe	Limited edition designs	I'm not

## 6.3. Creating database for InsideOut in SQL in Python

### Creating database & tables

```
5]: # Connect to SQLite database (or create it)
conn = sqlite3.connect('DEPIII_Projecttt.db')
cursor = conn.cursor()

# 1. Customers table
cursor.execute('''
CREATE TABLE IF NOT EXISTS customers (
    Customer_ID TEXT PRIMARY KEY,
    Gender TEXT,
    Location TEXT
);
''')

# 2. Products table
cursor.execute('''
CREATE TABLE IF NOT EXISTS products (
    Product_ID TEXT PRIMARY KEY,
    Product_Type TEXT
);
''')
```

#sql query

```
query = '''
SELECT DISTINCT
    platforms.Platform_name,
    countries.Country,
    countries.Percentage AS Country_Percentage
FROM
    countries
JOIN
    platforms ON countries.Platform_ID = platforms.Platform_ID
'''
Countries_sql = pd.read_sql_query(query, conn)
print(Countries_sql)
```

	Platform_name	Country	Country_Percentage
0	Instagram	Egypt	94.90%
1	Instagram	Saudi Arabia	1.00%
2	Instagram	Algeria	0.60%
3	Instagram	Morocco	0.30%
4	Instagram	Kuwait	0.30%
5	Facebook	Libya	1%
6	Facebook	Tunisia	1%
7	Facebook	Egypt	93%
8	Facebook	Saudi Arabia	1%
9	Facebook	Algeria	1%

## 6.4. Pre-processing

### 6.4.1. Sample of data summary in python (sales & marketing sheets):

#### Data Summary

```
# Marketing data summary

# summary for facebook (111)
df_marketing[df_marketing['Platform_ID'] == 111].describe()
```

	Date_ID	likes	comments	shares	views	Platform_ID
count	155.000000	155.000000	155.000000	155.000000	155.000000	155.0
mean	62668.400000	37.677419	4.258065	2.329032	1995.006452	111.0
std	38656.939045	121.463047	16.585019	8.419296	5173.264912	0.0
min	12023.000000	2.000000	0.000000	0.000000	125.000000	111.0
25%	32023.000000	10.000000	0.000000	0.000000	417.500000	111.0
50%	52024.000000	17.000000	1.000000	1.000000	752.000000	111.0
75%	92023.000000	29.500000	2.000000	2.000000	1555.000000	111.0
max	122023.000000	1225.000000	182.000000	90.000000	45800.000000	111.0

### 6.4.2. Sample of data cleaning in python (sales & marketing sheets):

#### Filling Missing values

```
10]: #checking missing values in customers data
df_customers.isnull().sum()
```

```
10]: Customer_ID      0
      Gender         0
      Location       12
      dtype: int64
```

```
11]: #checking missing values in orders data
df_orders.isnull().sum()
```

```
11]: Date_ID          0
      Order_ID        0
      Customer_ID     0
      Product_ID      0
      Order_Status    0
      No_of_products  50
      Product cost    0
```

#### Wrong format (changing data type)

```
45]: # change no_of_products data type from float to int
df_orders['No_of_products'] = df_orders['No_of_products'].astype(int)
```

```
# checking data type
print(df_orders["No_of_products"].dtype)
```

```
int32
```

```
46]: # Convert the date column from object to datetime
df_date['Date'] = pd.to_datetime(df_date['Date'], format='%m/%d/%Y')
# Check the data types to confirm the conversion
print(df_date["Date"].dtypes)
```

```
datetime64[ns]
```

#### 6.4.3. Sample of data summary in R (sales & marketing sheets):

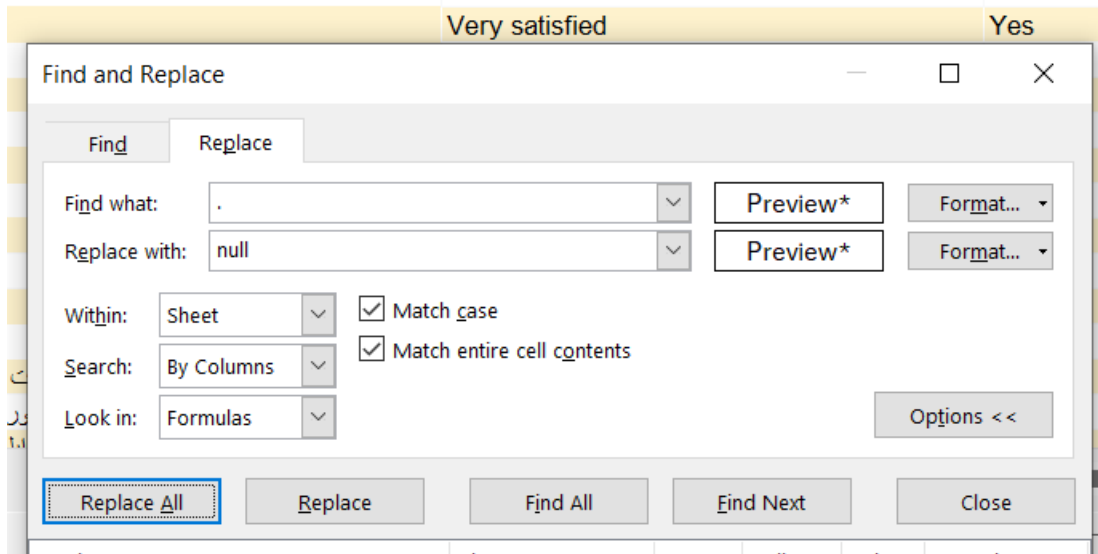
```
269 #wrong format
270 #changing data type
271 # Change 'No_of_products' column into Orders table to integer
272 df_Orders$No_of_products <- as.integer(df_Orders$No_of_products)
273
274 # Check data type of 'No_of_products'
275 str(df_Orders$No_of_products)
276
277 # Convert the 'Date' column into Date table to Date format
278 df_Date$Date <- as.Date(df_Date$Date, format = "%m/%d/%Y")
279
280 # Check the data type of the 'Date' column
281 str(df_Date$Date)
282 #Change percentage data type into Gender table |
```

#### 6.4.4. Sample of data cleaning in python (survey):

```
[17]: del survey['Timestamp']
```

```
[16]: del survey['Age_Group']
      del survey['Gender']
```

#### 6.4.5. Sample of data cleaning in Excel (survey):



#### 6.4.6. Sample of data cleaning in SQL (survey):

```

4  -- data cleaning
5  -- delete time column
6  • ALTER TABLE `surveyy`.`customerfeedback`
7    DROP COLUMN `i»Timestamp`;
8
9  -- delete age group
10
11 • ALTER TABLE `surveyy`.`customerfeedback`
12   DROP COLUMN `Age_Group`;
13   -- delete gender
14 • ALTER TABLE `surveyy`.`customerfeedback`
15   DROP COLUMN `Gender`;
16

```

### 6.5. Exploratory Data Analysis

#### 6.5.1. Sample of data analysis in excel (sales & marketing sheets):

Row Labels	Sum of No. of Products
Jan	131
Feb	81
Mar	49
Apr	81
May	111
Jun	54
Jul	31
Dec	228
<b>Grand Total</b>	<b>766</b>

6.5.2. Sample of data analysis in excel (Survey sheet):

Location Representation	
Row Labels	Count of Location
Cairo	105
Delta	55
Giza	41
Alexandria	22
Upper Egypt	21
<b>Grand Total</b>	<b>244</b>

6.5.3. Sample of data analysis in SQL (sales & marketing sheets):

```

SELECT
    platforms.Platform_name,
    marketing.Content_type,
    marketing.likes,
    marketing.comments,
    marketing.shares,
    marketing.views
FROM
    marketing
JOIN
    platforms ON marketing.Platform_ID = platforms.Platform_ID
...

```

6.5.4 Sample of data analysis in SQL (Survey sheet):

```
SELECT 'Planner Accessories' AS product, COUNT(*) AS count
FROM `surveyy`.`customerfeedback`
WHERE Interested_Products LIKE '%Planner accessories%'
UNION ALL
SELECT 'Notebooks' AS product, COUNT(*) AS count
FROM `surveyy`.`customerfeedback`
WHERE Interested_Products LIKE '%Notebooks%'
UNION ALL
SELECT 'To-do Lists' AS product, COUNT(*) AS count
FROM `surveyy`.`customerfeedback`
WHERE Interested_Products LIKE '%To-do lists%'
UNION ALL
SELECT 'Memos' AS product, COUNT(*) AS count
```

### 6.5.5. Sample of data analysis in python (sales & marketing sheets):

```
[79]: # pivoting percentage values for cities for each platform

Cities_Pivot = Cities_sql.pivot_table(index='City', columns='Platform_name', values='City_Percentage', aggfunc='sum')
Cities_Pivot
```

```
[79]:
```

	City_Percentage	
Platform_name	Facebook	Instagram
City		
Alex	9%	10%
Cairo	21%	40%

### 6.5.6 Sample of data analysis in Python (Survey sheet):

```
problem = survey.pivot_table(index='Satisfaction', values='Encountered_Problems', aggfunc='count')
Encountered_Problems = survey[survey['Encountered_Problems'] == 'Yes']

# Count occurrences of 'Encountered Problems' by Satisfaction
Encountered_Problems_grouped = Encountered_Problems.pivot_table(index='Satisfaction', values='Encountered_Problems', aggfunc='count')
Encountered_Problems_grouped
```

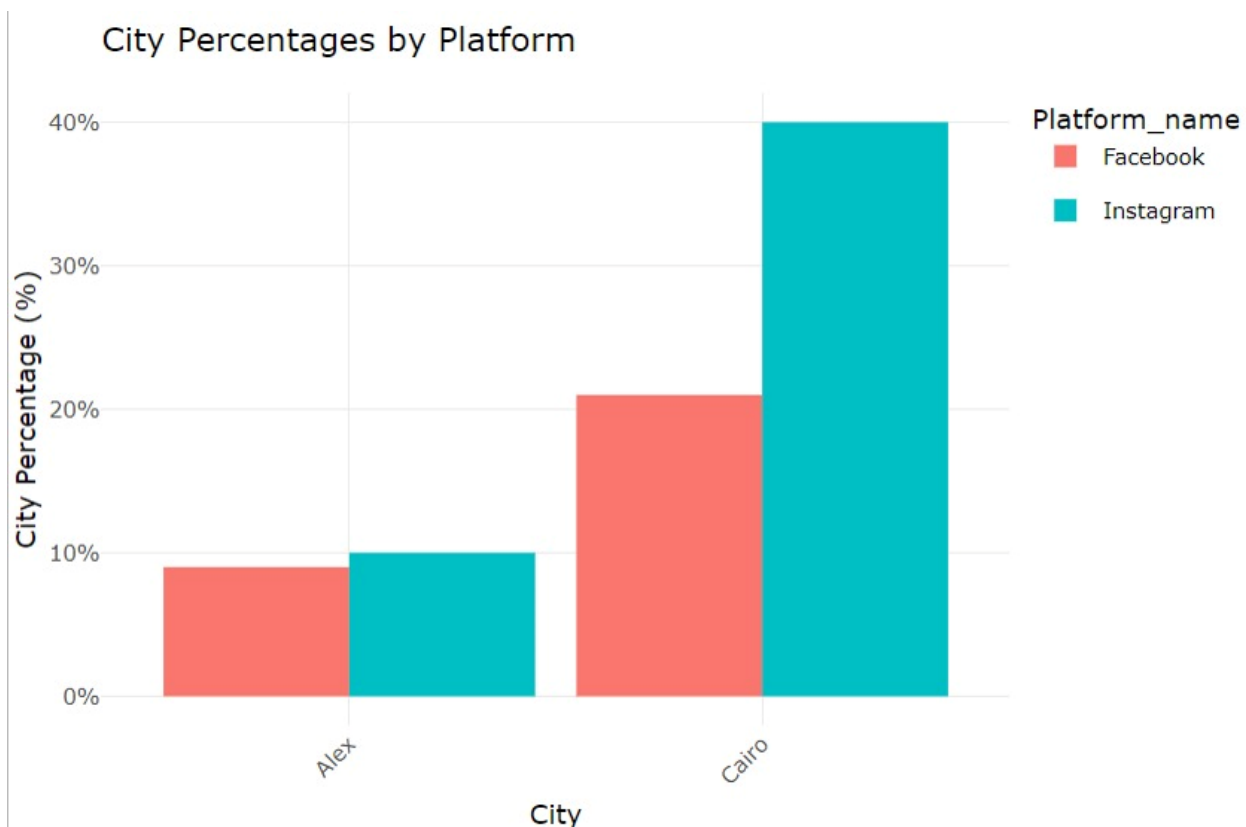
```
Encountered_Problems
```

Satisfaction	
Satisfied	13
Very satisfied	14

### 6.5.7. Sample of data analysis in R (sales & marketing sheets):

```

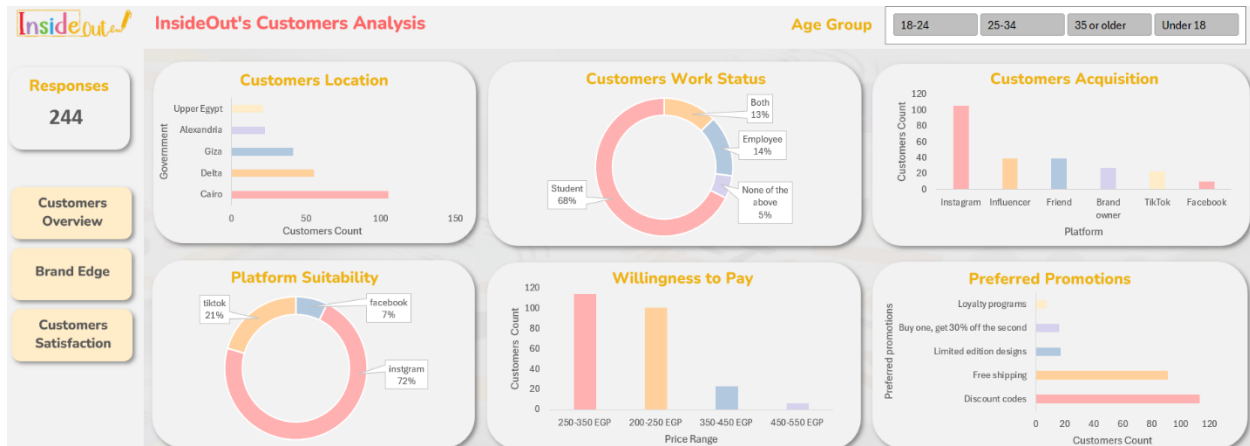
705
706 # Sample data (based on the data you provided)
707 cities_sql <- data.frame(
708   Platform_name = c("Instagram", "Instagram", "Facebook", "Facebook"),
709   City = c("Cairo", "Alex", "Cairo", "Alex"),
710   city_Percentage = c("40%", "10%", "21%", "9%")
711 )
712
713 # Clean the 'city_Percentage' column by removing '%' and converting to numeric
714 cities_sql <- cities_sql %>%
715   mutate(city_Percentage = as.numeric(gsub("%", "", city_Percentage)))
716
717 # Create a ggplot object to visualize city percentages by platform
718 plotly_bar <- ggplot(cities_sql, aes(x = City, y = city_Percentage, fill = Platform_n
719   geom_bar(stat = "identity", position = "dodge") +
  
```



## 6.7. Data Visualization

### 6.7.1. Data Visualization in Excel





## 6. 7.2. Data Visualization in Tableau

