📄 **DATA ANALYST INTERN ASSIGNMENT REPORT**

1️⃣ Introduction

This report presents an analysis of customer engagement and purchasing behavior using three datasets provided as part of the **Data Analyst Intern Assignment**. The objective is to clean and prepare the data, analyze the **user engagement → purchase funnel**, and derive meaningful **business insights** using data analysis and visualization techniques.

2️⃣ Datasets Description

The analysis uses the following datasets covering a two-week period:

2.1 Customers Dataset

- customer_id – Unique identifier for each customer
- registered_date – Date when the customer registered
- city – Customer's city
- acq_channel / acquisition_channel – Customer acquisition source

2.2 Orders Dataset

- order_id – Unique order identifier
- customer_id – Customer placing the order
- order_date – Date of order
- order_status – Status of the order
- net_sales – Order revenue
- disc% – Discount applied

2.3 Events Dataset

- user_id – User identifier
- event_date – Event date in YYYYMMDD format
- event_name – Type of user event

3️⃣ Data Cleaning & Preparation

The following steps were performed to prepare the data for analysis:

1. Standardized all column names to lowercase and underscore format

2. Converted date columns (registered_date, order_date, event_date) to proper datetime format

3. Removed duplicate records from all datasets

4. Renamed user_id to customer_id for consistency

5. Ensured customer_id data type consistency across datasets

6. Filtered valid orders using commonly accepted successful order statuses

7. Handled missing and invalid values safely to prevent runtime errors

These steps ensured the datasets were clean, consistent, and ready for analysis.

4️⃣ User Engagement → Purchase Funnel Analysis

4.1 Funnel Metrics

The funnel analysis evaluates how user engagement translates into purchases.

- **Active Users**: Unique users who triggered at least one event

- **Ordering Users**: Unique users who placed at least one valid order

Conversion Rate Formula:

Conversion Rate = Ordering Users ÷ Active Users

4.2 Funnel Results

- Active Users: **147,336**

- Ordering Users: *Calculated after valid order filtering*

- Conversion Rate: *Derived using the above formula*

5️⃣ Daily Trend Analysis

A time-series analysis was performed to observe daily user behavior.

Visualization Includes:

- Daily Active Users

- Daily Ordering Customers

A line chart was created to compare engagement and ordering trends over time. Defensive checks were implemented to avoid plotting when no valid order data was available.

6️⃣ Customer Behavior Insight

To understand customer purchasing behavior:

- Orders data was joined with customer data using customer_id

- Revenue was analyzed by **acquisition channel**

- A bar chart was used to visualize total revenue by channel

Key Insight:

The acquisition channel generating the highest total revenue contributes most significantly to overall sales.
Customers acquired through this channel demonstrate higher purchasing value.
Focusing marketing efforts on this channel could help maximize revenue.

7️⃣ Assumptions

The following assumptions were made during analysis:

- Since the dataset did not consistently label successful orders as "valid", commonly accepted successful statuses (completed, success, delivered) were treated as valid orders.

- Invalid or missing date values were ignored using safe coercion methods.

All assumptions were clearly stated to maintain transparency.

8️⃣ Tools & Technologies Used

- Python

- Pandas

- Matplotlib

- Jupyter Notebook / Google Colab

🔟 Conclusion

This project demonstrates a complete data analytics workflow including data cleaning, exploratory analysis, visualization, and insight generation.
The analysis follows best practices and provides actionable business insights, making it suitable for both real-world decision-making and internship evaluation