

# The Manhattan Project

presents

## CHALLENGE 01



# Briefing Time - 8pm tonight

## CHALLENGE 01

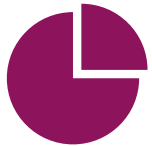




Problem Statement



Artifact  
Submission



Recommendation



Data Description



Challenge Evaluation  
Criteria



Vibelines  
& Bounties



# Problem Statement

## The Retail Challenge: Optimizing Customer Experience and Sales at *OmniMart Retailers*

### Problem Statement:

OmniMart Retailers is a multinational company with a vast database of customer transactions and feedback. The company's goal is to gain a deeper understanding of its customer base to improve sales, increase customer retention, and optimize its marketing strategies.

Your challenge is to act as a data analyst for OmniMart. Using the provided dataset, your team must perform a comprehensive Exploratory Data Analysis (EDA) to uncover actionable insights. Your analysis should focus on answering key business questions and identifying opportunities for growth.

### Your mission, is to:

#### 1. Analyze Customer Behavior:

Identify distinct customer segments based on their purchasing habits, demographics, and income.

Determine the average amount spent per customer and analyze how spending varies across different customer segments (e.g., age, gender, location).

Uncover trends in customer loyalty by examining the frequency of purchases and the total number of purchases made.

#### 2. Evaluate Product Performance:

Identify the most popular and least popular product categories and brands.

Analyze the relationship between product feedback (ratings) and sales performance.

Investigate if certain product types are more popular with specific customer segments.

#### 3. Optimize Operations and Logistics:

Analyze the impact of different shipping methods on customer order status (e.g., delivered vs. canceled).

Examine the most common methods of payment and their correlation with order value.

#### 4. Actionable Insights & Recommendations:

Based on your findings, prepare a detailed report or presentation that outlines key insights and provides clear, data-driven recommendations to the OmniMart leadership team.

Think about how your insights can be used to create targeted marketing campaigns, improve the product catalog, or enhance the overall customer experience.



# Data Description

Certainly! The dataset you're referring to contains detailed information about retail transactions in *OmniMart Retailers* aimed at providing a comprehensive view of customer behavior and purchasing patterns. Here's a breakdown of the various attributes and what they entail:

## Customer Information:

Customer ID: A unique identifier for each customer.

Name: Customer's full name.

Email: Customer's email address for communication.

Phone: Contact number of the customer.

Address: Physical address of the customer.

City, State, Zipcode, Country: Geographic details of the customer.

Age: Age of the customer.

Gender: Gender of the customer.

Income: Income bracket or level of the customer.

Customer Segment: Classification of customers based on behavior or demographics.

## Transaction Details:

Last Purchase Date: Date of the customer's most recent purchase.

Total Purchases: Total number of purchases made by the customer.

Amount Spent: Total monetary amount spent by the customer.

## Product Information:

Product Category: Category to which the purchased product belongs (e.g., electronics, clothing, groceries).

Product Brand: Brand name of the product.

Product Type: Type or model of the product purchased.

## Feedback:

Feedback: Customer's feedback or rating related to the product or service received.

## Transaction Logistics:

Shipping Method: The method used to deliver the purchased products.

Payment Method: The method of payment chosen by the customer.

Order Status: Status of the order (e.g., shipped, delivered, canceled).



# Artifact Submission

Your team's submission must include the following five artifacts, all packaged within a single GitHub repository.

## 1. Jupyter Notebook (.ipynb)

This is the core of your submission. Your Jupyter Notebook should be a complete, well-documented narrative of your data analysis journey. It must include:

- **Detailed Explanations:** Use Markdown cells to explain your thought process, the questions you are trying to answer, and the insights you've uncovered.
- **Clean Code:** The code should be well-structured, easy to read, and free of unnecessary clutter.
- **Comprehensive Comments:** Use comments to explain complex logic and the purpose of different code blocks.
- **Key Visualizations:** All visualizations should be clear, properly labeled, and directly support your findings.

## 2. Presentation (.pptx or .pdf)

Create a compelling presentation (10-15 slides) that summarizes your team's analysis and key findings. This presentation should serve as your final pitch to the judges. It must include:

- **Executive Summary:** A concise overview of your findings.
- **Key Insights:** The most important takeaways from your analysis.
- **Data-Driven Recommendations:** Actionable steps that OmniMart can take based on your insights.
- **Supporting Visualizations:** A selection of your best visualizations to illustrate your points.

## 3. README File (.md)

The README file is the first thing we'll look at. It should serve as a quick guide to your project and provide essential details. It must include:

- **Project Title & Team Members:** Your team's name and the names of all members.
- **Brief Problem Statement:** A summary of the challenge and your approach.
- **Summary of Findings:** A bullet-point summary of your most significant insights.

## 4. Attached Dataset

Please include the original dataset (.csv or other format) within your repository. This ensures the judges can reproduce your analysis without any issues.

## 5. GitHub Repository

Your final submission will be your GitHub repository. The repository name **must follow this exact format:**

**Teamname\_Challenge1\_TMP**



# Challenge Evaluation Criteria

Criteria	Weightage	Expectation
1. Data Understanding and Initial Insights	30%	<ol style="list-style-type: none"><li>1. Summarized the dataset's overall structure, purpose, and key variables to establish clear context.</li><li>2. Generated essential summary statistics and visualizations to describe the distribution and core characteristics of the data.</li></ol>
2. Visualization and Pattern Identification	25%	<ol style="list-style-type: none"><li>1. Selected and created appropriate visualizations (e.g., histograms, scatter plots, box plots) to effectively explore the distribution and relationships of variables.</li><li>2. Ensured all visualizations are clear, well-labeled, and easy to interpret, effectively communicating the intended message without ambiguity.</li><li>3. Identified and highlighted meaningful patterns, trends, correlations, and outliers directly from the visualizations.</li><li>4. Employed a variety of visualization techniques to analyze both univariate and bivariate relationships within the dataset.</li><li>5. Summarized the key insights from the visualizations in a clear narrative, directly linking these findings to the project's overall objectives.</li></ol>
3. Statistical Analysis and Hypothesis Generation	20%	<ol style="list-style-type: none"><li>1. Utilized descriptive statistics to summarize the dataset's key features, including measures of central tendency (e.g., mean, median, mode) and dispersion (e.g., standard deviation, variance, quartiles).</li><li>2. Identified and explained relationships between variables using statistical metrics, such as correlation coefficients, to inform subsequent analysis.</li><li>3. Formulated clear and testable hypotheses based on initial data insights and descriptive statistics.</li><li>4. Summarized the findings of the statistical analysis in a clear narrative, drawing conclusions that contribute to a deeper understanding of the data.</li></ol>
4. Data Storytelling of Findings and Recommendation	15%	<ol style="list-style-type: none"><li>1. Structured the analysis as a clear, compelling narrative, connecting the initial problem statement to the final insights and recommendations.</li><li>2. Translated complex technical findings and visualizations into a clear, non-technical narrative that is easy for a business audience to understand.</li><li>3. Formulated specific, data-driven recommendations that are actionable and directly address the initial problem.</li><li>4. Quantified or clearly articulated the potential business impact and value of the recommendations.</li><li>5. Backed all findings and recommendations with strong, compelling evidence from the data and visualizations presented.</li></ol>
5. Coding Guidelines and Standards	10%	<ol style="list-style-type: none"><li>1. Code is well-structured, organized, and easy to follow.</li><li>2. Meaningful variable names and comments are used throughout the code.</li><li>3. Code is modular and functions are used effectively to promote reusability.</li><li>4. Adhered to consistent coding style conventions (e.g., PEP 8 for Python).</li><li>5. Project includes clear documentation (e.g., README file) outlining the project goals, data sources, steps taken, and how to run the code.</li></ol>





# Recommendation for the Notebook

## Phase 1: Preparation & Setup

**Project Initialization:** Import all necessary Python libraries (e.g., pandas, numpy, matplotlib, seaborn).

**Data Acquisition:** Load the challenge dataset into your notebook.

**Data Scrutiny (The Cleanup Mission):** Begin your data cleaning. Identify and remove any "missing" values to ensure your analysis is based on clean, reliable data.

## Phase 2: Initial Reconnaissance

**Initial Survey:** Get to know your data. Print the schema (.info()), a statistical summary (.describe()), the shape of the dataset (.shape), and a preview of the top rows (.head()).

**Data Classification:** Segregate all variables into their logical types: numerical, categorical, and temporal. This will guide all your future analyses.

## Phase 3: Uncovering Insights (The Core Mission)

**Univariate Analysis:** Dive into each variable individually.

For **numerical** features, use histograms, box plots, and density plots to understand their distribution.

For **categorical** features, use bar charts and count plots to see the frequency of each category.

**Bivariate Analysis:** Start exploring relationships between two variables.

**Numerical vs. Numerical:** Use scatter plots and correlation matrices to identify relationships (e.g., does energy correlate with loudness?).

**Numerical vs. Categorical:** Use box plots and violin plots to compare the distribution of a numerical variable across different categories (e.g., how does the duration\_ms vary by language?).

**Multivariate Analysis:** Now, look at three or more variables at once. This is where you can create complex, insightful visualizations (e.g., a scatter plot of popularity vs. danceability, with different colors for each language).

## Phase 4: Special Operations

**Temporal Analysis:** Use the year feature to analyze trends over time. How has the popularity or danceability of songs changed over the years?

**Outlier Analysis :** Use statistical methods like the Interquartile Range (IQR) to identify and analyze outliers, specifically using popularity as the key variable. Investigate what makes these outliers so unique.

## Phase 5: The Final Briefing

**Actionable Recommendations:** Based on all your findings, synthesize your insights into concrete, data-driven recommendations for the mixing engineer. Back up every recommendation with the evidence you found in your analysis.



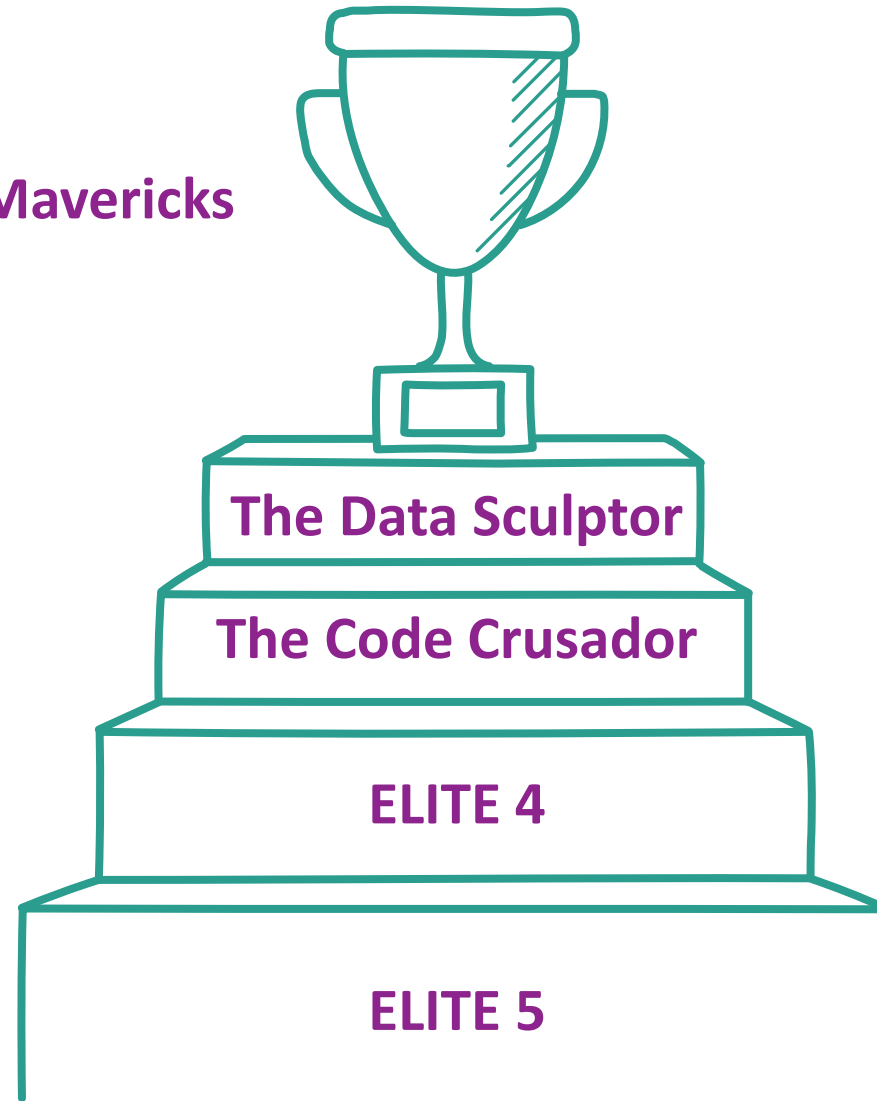


# VIBELINES

Kickoff Vibes - 12<sup>th</sup> Sept  
Wrapoff Vibes - 14<sup>th</sup> Sept  
Spotlight Vibes - 16<sup>th</sup> Sept



The Data Mavericks



5k



3k



2k



# Important Links

## Challenge Brief and Dataset

[https://github.com/manhattanproject-ai/Challenge1\\_TMP\\_OmniMart-Retailers/tree/main](https://github.com/manhattanproject-ai/Challenge1_TMP_OmniMart-Retailers/tree/main)

## Submission Link

<https://docs.google.com/spreadsheets/d/1KNNVZJBU8LbMTYb-HUnvEFPpbdxFvtQgVHJ74EG21I0/edit?pli=1&gid=557188903#gid=557188903>



# Lets Go

