# Skillspire Marketing Analytics Challenge

\*\*Decode Customer Behavior with Data\*\*

## Overview

You’ve just been hired as a Marketing Data Analyst at RetailNova, a rapidly growing e-commerce company. The leadership team wants to understand customer behavior, evaluate marketing campaign performance, and identify the factors driving sales and churn. Your task: Use Python-based analytics to uncover insights, visualize trends, and build models that could help RetailNova optimize its marketing strategy.

## Dataset Description

The dataset (retailnova\_marketing.csv) contains 10,000 simulated customer records including demographic, behavioral, and campaign data. Each row represents a unique customer.

|  |  |
| --- | --- |
| Variable | Description |
| customer\_id | Unique customer identifier |
| age | Customer age (18–70) |
| gender | Male / Female / Other |
| region | North, South, East, West |
| acquisition\_channel | Customer acquisition source (Social Media, Search Ads, Referral, Email, Organic) |
| signup\_date | Date of account creation |
| last\_purchase\_date | Date of most recent purchase |
| total\_purchases | Number of purchases in the past 12 months |
| avg\_order\_value | Average order value in USD |
| campaign\_exposure | Whether the customer saw the recent ad campaign (Yes/No) |
| campaign\_clicks | Number of ad clicks |
| time\_on\_site | Average time spent on website (minutes per session) |
| churned | 1 if customer hasn’t purchased in 6 months, else 0 |
| revenue\_last\_12m | Total revenue from the customer in the last 12 months |

## Challenge Tasks

### Part 1 — Descriptive Statistics (30 points)

• Summarize customer demographics (age, gender, region)  
• Calculate overall purchase frequency and average order value  
• Identify the top-performing acquisition channels  
• Compare engagement metrics (time on site, clicks) between churned vs active users  
• Provide short written insights for each finding

### Part 2 — Data Visualization (30 points)

Create Python-based visuals using matplotlib, seaborn, or plotly:  
• Distribution of customer ages and purchase frequency  
• Monthly sales trend or campaign performance over time  
• Relationship between campaign clicks and total purchases  
• Churn rate by region or acquisition channel  
  
Bonus: Create an interactive dashboard with Plotly or Streamlit.

### Part 3 — Statistical Modelling (40 points)

• Hypothesis Testing: Assess if campaign exposure significantly increases total purchases  
• Test if average order value differs across acquisition channels  
• Build a classification model (logistic regression or random forest) to predict churn probability  
• Identify top predictors and evaluate model accuracy

## Scoring Rubric

|  |  |  |
| --- | --- | --- |
| Category | Criteria | Points |
| Data Exploration | Cleaning, completeness, descriptive analysis | 20 |
| Visualization & Storytelling | Quality, clarity, interpretability of visuals | 30 |
| Modeling & Analysis | Model accuracy, justification, interpretability | 30 |
| Insight & Business Impact | Depth of insights and marketing implications | 15 |
| Creativity / Bonus | Innovative analysis or presentation | 5 |

## Deliverables

Participants must submit a Jupyter Notebook (.ipynb) containing:  
• Clean, well-commented Python code  
• At least three visualizations  
• Statistical results and written interpretations  
• Final insights and recommendations  
  
Optional: HTML or PDF export of the notebook for presentation.

## Submission Tips

• Document your thought process and assumptions clearly.  
• Focus on actionable marketing insights, not just technical correctness.  
• Use data storytelling to communicate findings to non-technical audiences.  
• Creativity counts — think like a data-driven strategist, not just a coder.

## Example Insights

• Customers acquired via Social Media have higher engagement but lower retention.  
• Campaign exposure increased total purchases by 12%, with statistical significance (p < 0.05).  
• Logistic model shows time\_on\_site and total\_purchases are strong predictors of churn.