Achievement 4 Project

Regional Segmentation and Customer Profiling Instacart Grocery Basket Analysis

Objective

- The objective of this project is to conduct a comprehensive analysis of Instacart's grocery basket data to gain insights into customer behavior and preferences across different regions. The project aims to achieve the following:
- **Regional Segmentation:** Identify and categorize customers based on their geographic location to understand regional differences in shopping habits and spending patterns.
- **Customer Profiling:** Develop detailed customer profiles by analyzing demographic factors such as age and income. This profiling will help in understanding the distinct characteristics of various customer groups and their impact on purchasing behavior.
- **Spending Analysis:** Examine and compare average spending across different customer profiles and regions to identify key trends and potential opportunities for targeted marketing strategies.
- **Exclusion of Low-Activity Customers:** Filter out low-activity customers to focus the analysis on more engaged users, thereby providing a more accurate representation of active customer segments.
- **Data Visualization:** Create visual representations of the findings to facilitate better understanding and communication of insights derived from the data.
- The ultimate goal is to provide actionable insights that can help Instacart optimize its marketing efforts, improve customer retention, and increase overall sales.

Context

- In the rapidly evolving grocery delivery market, understanding customer behavior and preferences is crucial for companies like Instacart to remain competitive. Instacart, a leading online grocery platform, connects customers with local stores, enabling them to order groceries and have them delivered to their doorstep. As the platform continues to grow, the vast amount of data generated from customer interactions offers valuable opportunities for analysis and strategic decision-making.
- This project focuses on analyzing Instacart's grocery basket data to uncover patterns and trends in customer purchasing behavior. By segmenting customers based on geographic regions and profiling them according to demographic characteristics, this analysis aims to provide deeper insights into how different customer groups interact with the platform. Understanding these nuances is essential for optimizing marketing strategies, personalizing customer experiences, and ultimately driving growth.
- The project also addresses the challenge of managing and analyzing large datasets, ensuring that only the most relevant and active customer segments are considered in the analysis. Through a combination of data processing, statistical analysis, and visualization, this project seeks to deliver clear and actionable insights that can inform Instacart's business decisions and enhance its competitive edge in the market.

Key Questions

This project seeks to answer the following key questions to drive actionable insights from Instacart's grocery basket data:

1.How do customer spending patterns vary across different regions?

- •What are the average spending levels in each geographic region?
- •Are there significant differences in purchasing behavior between regions?

2.What distinct customer profiles can be identified based on demographic factors such as age and income?

- •How do different age groups and income levels influence purchasing habits?
- •What are the characteristics of the most active customer segments?

3.Which regions and customer profiles show the highest potential for targeted marketing strategies?

- •How can regional differences and customer profiles be leveraged to tailor marketing campaigns?
- •Which customer segments demonstrate the highest levels of engagement and spending?

4.How does excluding low-activity customers impact the overall analysis?

- •What changes are observed in spending patterns and customer profiles when low-activity customers are excluded?
- •Does focusing on active customers provide more accurate and actionable insights?

5.What visualizations can effectively communicate the insights derived from the analysis?

- •Which charts and graphs best represent the distribution of customer profiles across regions?
- •How can visualizations be used to highlight key trends and differences in customer behavior?

By addressing these questions, the project aims to provide a comprehensive understanding of Instacart's customer base, identify opportunities for growth, and support data-driven decision-making in marketing and customer relationship management.

Stakeholders

- **Vice President of Marketing:** “We’re always looking to improve our targeting for ad campaigns. Understanding customer behavior and preferences will help us deliver more personalized and effective marketing strategies.”
- **Instacart Customer:** “I want to receive ads, promotions, and recommendations that are relevant to the products I order regularly. Personalized communication enhances my shopping experience.”

- **Senior Vice President of Sales:** “We need to identify which parts of our offering have the lowest market share and understand why. This insight will allow us to improve these areas and boost overall sales.”

Data

- Throughout this Achievement, you'll be using a number of open-source data sets from
- Instacart. You'll also receive a customer data set (created and included for the purpose of
- this project), on which you'll apply what you've learned to address the project's key
- questions. While each data set contains a different kind of information, they all include some
- kind of common identifier.
- The project data you'll need is linked for reference below. However, you'll receive links to
- each data set in the Exercise content, as well.

Career Foundry Data Sets:

- Customers Data Set

•Instacart Data Sets:

- ● Data Dictionary
- ● Citation (required in your final report): “The Instacart Online Grocery Shopping
- Dataset 2017”, Accessed from www.instacart.com/datasets/grocery-shopping-2017
- via Kaggle on <date>.

Analysis Criteria

Project folder follows industry standards in terms of structure and naming

- **conventions.**
- **● Analysis has been conducted using Jupyter notebooks and the Anaconda libraries**
- **manager.**
- **● Analysis has been conducted using Python and relevant libraries (pandas, NumPy, os,**
- **matplotlib, scipy, and seaborn).**
- **● All required libraries have been successfully installed and imported into each script.**
- **● Python scripts are clean and easy to follow with headings and contents lists.**
- **● All code is consistent (e.g., with the use of quotation marks and spaces) and includes**
- **descriptive comm ents.**
- **● All required data sets have been successfully installed and imported into each script.**
- **● Descriptive checks have been conducted after importation of data, such as checking**
- **the top and the bottom of the dataframe.**
- **● Whenever a dataframe is altered, checks have been conducted to determine its**
- **shape and basic statistics.**
- **● All project data has been merged into a single data set. A frequency of the merge flag**
- **shows the merged data set is a 100% match to the combined original data set‘s**

Analysis Criteria

- Merged data set only contains variables to be used in the analysis.
- ● All column names are self-explanatory.
- ● All identifier variables follow the industry standard data type.
- ● Data has been cleaned. Duplicate data, missing data, and mixed-type columns have
- been checked and addressed.
- ● Samples have been exported whenever an exclusions flag has been created.
- ● All subsamples have been exported and saved in the proper folder following a
- consistent naming convention.
- ● Any new columns that have been derived are relevant to the needs of the analysis.
- ● At least 4 types of data visualizations have been generated to communicate insights
- to stakeholders. Visualizations are clearly labeled.
- ● Data ethics have been kept in mind when dealing with data, especially in regards to
- customer information.
- ● Final report includes evidence of analysis methodology, clear answers to the
- questions in this brief, and recommendations for Instacart stakeholders.
- ● Final report contains data citation for Instacart and customer data sets.

Terminology

- In analytics, a single procedure or concept can often be called a variety of different things.
- We've aimed to be consistent with the terminology used throughout this Achievement. Even
- so, there are a few variations that come up again and again when conducting an analysis in
- Python. We've included a list below to help you navigate this terminology:
- **● script = notebook**
- **● variable = column = characteristic**
- **● observation = entry**
- **● dataset = dataframe = df**
- **● read = import**
- **● run = execute**
- **● write = export = save**
- **● derive a variable = create a column**
- **● filter = subset**
- **● merge flag = match flag**
- **● key column = identifier column**

Project Tasks & Deliverables

- Throughout this Achievement, you'll work on your project from one Exercise to the next,
- completing tasks as you go. For each task, you'll submit a deliverable that makes up a piece
- of your project. Below is a breakdown of your tasks and deliverables by Exercise:

Chapter‘s

- **Exercise 1: Intro to Programming for Data Analysts**
- ● Install Anaconda
- ● Launch Jupyter
- **Exercise 2: Jupyter Fundamentals & Python Data Types**
- ● Create project folder
- ● Install required Python libraries
- ● Create a notebook and import libraries
- ● Practice coding using basic Python data types
- **Exercise 3: Introduction to Pandas**
- ● Download data and import into notebook as a pandas dataframe
- ● Conduct basic descriptive exploratory tasks
- **Exercise 4: Data Wrangling & Subsetting**
- ● Change data types of identifier variables into more suitable types and rename
- columns where needed
- ● Access values and determine their meaning using a data dictionary
- ● Create new dataframes based on a certain criteria
- ● Answer questions about user activities based on variable frequencies
- **Exercise 5: Data Consistency Checks**
- ● Fix mixed-type variables
- ● Uncover and deal with missing values
- ● Uncover and remove duplicates
- *Note: Instacart is a real company that’s made their data available online. However, the contents of this project brief*

Chapter‘s

- **Exercise 6: Combining & Exporting Data**
- ● Merge a set of given dataframes
- ● Analyze results from merge flag frequencies
- ● Export merged data as a pickle file
- **Exercise 7: Deriving New Variables**
- ● Create new columns using conditional logic in the form of if-statements, user-defined
- functions, the loc() function, and for-loops
- **Exercise 8: Grouping & Aggregating Data**
- ● Create flags, for instance, a loyalty flag, and place them in new columns
- ● Create summary columns of descriptive statistics using the groupby() function
- **Exercise 9: Intro to Data Visualization with Python**
- ● Import and prepare a customer data set
- ● Merge customer data with other project data
- ● Create histograms, bar charts, line charts, and scatterplots for different variables and
- relationships between variables
- **Exercise 10: Coding Etiquette & Excel Reporting**
- ● Create new columns and flags using customer data to inform customer profiling
- ● Analyze order behavior of different customer groups
- ● Summarize analysis findings and describe what connections in the data you’ve found
- ● Create a report that describes your analysis methodology, your results, and your
- recommendations for Instacart stakeholders