Segmentation-Based Credit Card Marketing Project for BluCard Credit

MGTA 457 Business Intelligence Systems : BI Project Document

Group 12

Srihari Nair, Yiying Zhou, Charlotte Wang, Diyansh Mankotia, Samruddhi Kulkarni

Introduction

We use the data of credit card transactions, US household community information, and US housing data to conduct a customer analysis on Credit Card Marketing Strategy. Our company, BluCard Credit Services, is trying to understand its customer base, thus promoting its new credit card products to each target group based on their specific needs. The dashboards are designed to unveil insights into customer preferences and behaviors, enabling BluCard to tailor its credit cards to different groups of consumers.

Project Goals

1. Build a customer analysis dashboard to understand the consumer behavior of consumers, and group consumers based on similar consumer profiles.
2. Understand the consumer spending behavior of the aimed categories, and design tailored credit cards for each consumer group.

Team Roles

- Initiation and Design Document: Srihari Nair, Yiying Zhou
- Data Research and Prep: Charlotte Wang
- Discussion and Creating Views: Srihari Nair, Yiying Zhou, Charlotte Wang, Samruddhi Kulkarni, Divyanshi Mankotia
- Creating Dashboard: Samruddhi Kulkarni, Divyanshi Mankotia

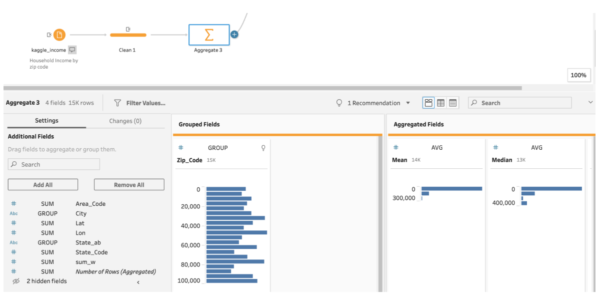
1. Data Source and Prep

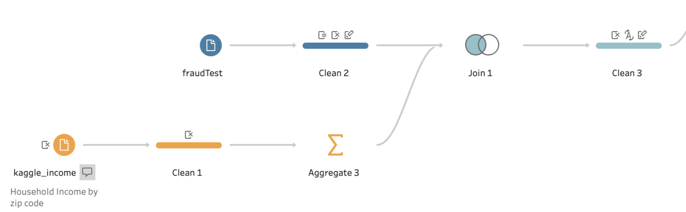
Datasets included in the project:

1) **Credit Card Transactions Dataset**
   https://www.kaggle.com/datasets/kartik2112/fraud-detection/data
   (Initial dataset, as proposed in the Design Document, contains credit card transaction details)
2) **Income:US Household Income Statistics**
   https://www.kaggle.com/datasets/goldenoakresearch/us-household-income-stats-geo-locations/
   (Contains household income information based on zip code within the U.S., to give more insights, trying to investigate any relations between community profile and spendings)
3) **Inventory: RDC Inventory Hotness Metrics Zip Historical Data**
   https://www.realtor.com/research/data/
   (Realtor.com data source, each zip code's housing price information, again to introduce more dimensions of the community's information and how they may relate to different kinds of customers)

Data Preparation and Cleaning steps:

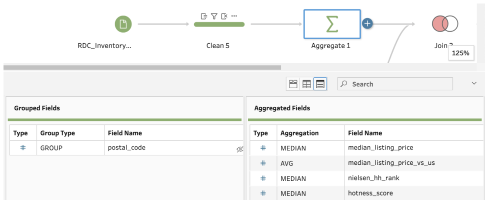
We began by examining the Credit Card Transaction Data, focusing on relevant fields and data types for potential modification or removal. Despite its vast number of records (over 300K), the dataset only spanned half of 2020. To depict longer-term trends, we extended the transaction dates to two years using Python before importing into Prep Builder. In Prep Builder, we first cleaned the fraudtest.csv file by eliminating the street address and converting birth dates to years for age categorization. Next, we refined the kaggle_income dataset, keeping only essential fields for regional analysis and removed excess location details. To address duplications in transaction details, we aggregated the dataset by zip code, averaging income-related fields.

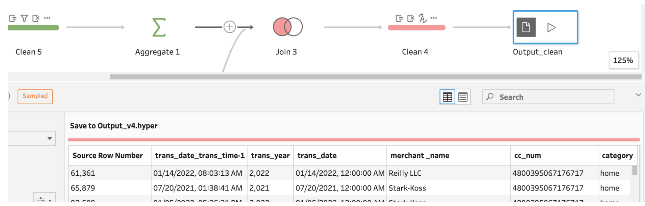




We performed a left join on the cleaned transaction and aggregated income datasets using "zipcode". After joining, we further refined the data by removing duplicate zip_code and row number fields, renaming hh_income columns, and converting credit card numbers from numeric to string format.

For the third dataset, RDC_Inventory (housing price data), we eliminated non-essential columns like supply and demand metrics, days on market, and monthly/yearly changes. We filtered for "high_quality" data, excluding rows flagged with a quality_flag of 1, and altered date granularity from daily to yearly. We then aggregated this dataset by zip code, averaging listing prices, community rankings, and hotness scores. Finally, we joined this aggregated housing data with the previous datasets using a left join.





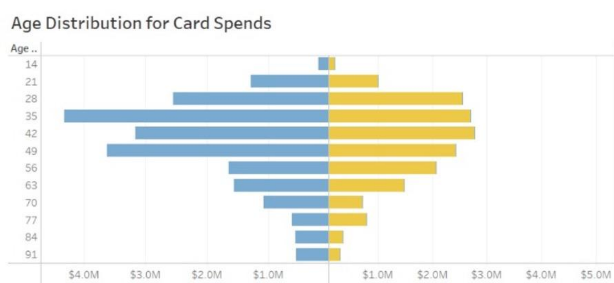
Before generating the final output, we applied the last and final cleaning step for the joined result from the 3 datasets. The transformations include chunking transaction dates, creating customer age, splitting merchant name, and removing fields that contain the same information as others. After all the steps being applied and output being inspected and making sense, we extracted the output file to a .hyper file to be ready to use for creating the views and dashboards.

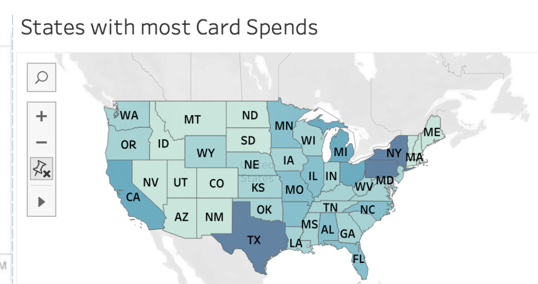
2. Dashboard Building

Dashboard One:

The role of the first dashboard is to explore the dataset and understand the customer base. We selected the average age, income, amount of spending and house listing price first, and found that our customers are mostly the middle aged high-income group, with an average of transaction amount.

To segregate our customers, we explored gender, age, state, job category, card spend, house listing price and purchase category. Based on these variables, we conducted five graphs to show the relationship between them.

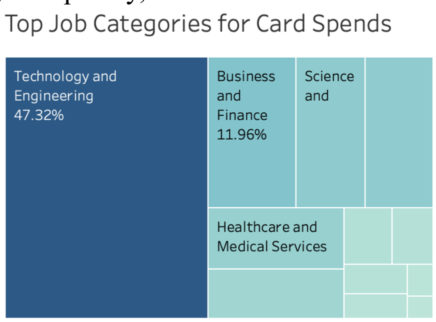


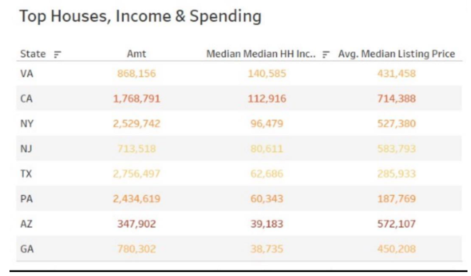


Based on age distribution and card spending we find that male and females in different age groups showed similar spending habits. For people between 28 to 49, this age group spends considerably more than others. For people under 25, they have lower spending amounts, and can be considered as potential target customers.

For geographic distribution, we can find that states like Texas, New York and California show higher card spendings compared to others. States like Utah, Arizona and Idaho show relatively low spendings. Therefore, our target groups for promoting the credit cards are those states with high spendings(darker blue color).

By aggregating the card spending based on groups, we revealed that each job group has different spending habits as well. Those working on technology, business and science showed high spending and high frequency, thus can be considered as a premium group for credit cards.





Top Houses, Income & Spending

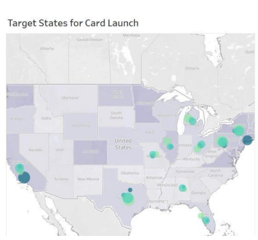
State	Amt	Median Median HH Inc..	Avg. Median Listing Price
VA	868,156	140,585	431,458
CA	1,768,791	112,916	714,388
NY	2,529,742	96,479	527,380
NJ	713,518	80,611	583,793
TX	2,756,497	62,686	285,933
PA	2,434,619	60,343	187,769
AZ	347,902	39,183	572,107
GA	780,302	38,735	450,208

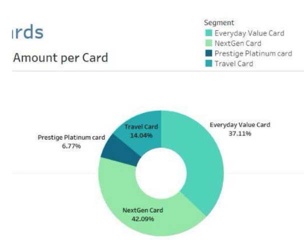
Analyzing housing prices, income, and credit card spending across states reveals varied credit card usage behaviors. Virginia, with high income but lower housing prices, exhibits medium credit card spending, identifying it as a potential market for promoting card usage. States like California, New York, and Texas, characterized by high incomes and spending, have more prestigious customers. Pennsylvania, with medium income and spending levels, represents frequent users. Conversely, Arizona and Georgia, despite high housing prices, show low spending, making credit card promotion challenging there.

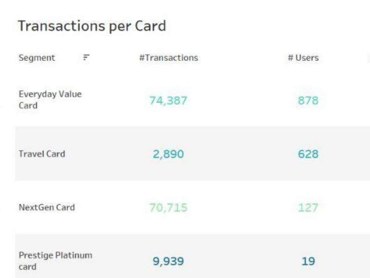
In terms of usage scenarios, high purchase amounts were noted in shopping and travel, while grocery purchases showed high frequency but lower amounts. This information will guide our targeting of credit card consumers based on purchase categories.

Dashboard Two:

Based on consumer portraits, we launched four types of credit cards:
Everyday Value: targeted based on shopping categories with " shopping", "grocery" ,"home" or "entertainment", with age above 25 years.
Prestige Value: Median housing income above 115,000 and median house listing price above 200,000, with states in Texas, California, New York, or Pennsylvania.
IsNextGen: Customer ages between 18 to 30.
IsTravel: Spending Category is travel with spending amount above 100.







Transactions per Card

Segment	#Transactions	# Users
Everyday Value Card	74,387	878
Travel Card	2,890	628
NextGen Card	70,715	127
Prestige Platinum card	9,939	19

Target states for card launch align with consumer behavior in terms of income and spending. Prestige Value card users typically have the highest income and property listing prices. IsTravel card holders, often not focused on property ownership, contribute to lower average property prices. NextGen Card and Everyday Value card dominate the market, with a large customer base and transaction volume. IsTravel and Prestige cards, though smaller in user numbers, generate significant transactions. Everyday Value cards are widely used frequently, whereas NextGen cards cater to a smaller group with high transaction needs. Travel card users form a large base but use their cards infrequently, and Prestige cardholders, although few, account for high transaction volumes.

3. Conclusion and Marketing Strategy

Based on the card types and their differences, we will launch different marketing strategies for these four groups:

- ➔ **Everyday Value card:** collaborate with grocery stores, retailers and/or shopping malls to promote our card. We will launch events such as "register credit card for free sample" or "register credit card for discount", and will be limited to states such as California, New York and Texas etc.
- ➔ **NextGen Card:** collaborate with universities to promote the NextGen Card. The NextGen Card will have the highest transaction amount limit since its user groups are mainly students, and also we will give special discounts when they are shopping online using NextGen Card.
- ➔ **IsTravel Card:** promote our IsTravel Card in airports and train stations. Users with this card can use it without state and country limits. Also users can have more discounts when they use the card for booking hotels, flying tickets and shopping in Duty-free shops, shops in the airports.
- ➔ **Prestige Card:** the card transaction will have less limitations, and users can enjoy special rewards when the transaction amount is above a certain benchmark. This card will also support the benefits of all the other types. We aim to increase the frequency and actual value spent for people who enjoy our premium card.

4. Reflection on Dashboard Design and Changes

In the design phases, since we don't have the cleaned dataset, we could only give a rough idea for what plot we wanted to use. And we only have one dataset initially, which was not sufficient.

After we start to build the dashboard, we enlarge the dataset we choose, combined with household and income data for analysis. Also, based on the data, line charts and heat maps are not used since they can't reveal any useful findings. Also, some parameters such as average amount of spending for each card type didn't show an ideal performance, so we removed it.

What we learned is that having a draft is important for building the final dashboard since it will help us understand what data we will need, how to combine different parameters and what plots may be used. Making agile changes in this process is also crucial. The final dashboard should clearly reveal all the parameters we want to show and the trends behind data.