Data Analysis Portfolio

Professional Skills

- Spreadsheets, Excel (Advanced)
- Structured Query Language(SQL)
- Data Visualization - Tableau
- Python - Statistical Programming
- Big Data - Google Cloud
- Detail - Oriented

Experience

I3 infosoft / Data Science Intern

March 2021 - May 2021, A-77, SECTOR-2, NOIDA

- Went through 3 months of extensive Industrial training on solving pre-existing problems.
- Collaborated with 5 other team members to accomplish Data extraction and mining using beautiful soup,
- Data validation and Data cleaning using Python code on tools like pandas, Numpy,
- Data visualization using matplotlib and seaborn libraries.
- Quickly Adapted to dynamic requirements to Build customized models to analyze and collect insights from the processed data.
- Presented our solutions in tailored web pages using Flask and deployed our models on AWS Cloud.

Education

J.C BOSE UNIVERSITY OF SCIENCE AND TECHNOLOGY

Bachelor of Technology in Computer Science AND Engineering

July 2017 - August 2021, FARIDABAD, INDIA

CGPA 7.71 / 10

Continuing Education

Google / Data Analytics Professional Certificate

June 2021 - August 2021, ONLINE

Continuing extensive six month job-ready Google Career Certificate Training. Demonstrated hands-on experience with data cleaning, data visualization, project management, interpreting and communicating data analytics findings. Confidence in transforming complex data into actionable and clear insights. Fluency in computer programming languages and a solid understanding of databases.

Professional Background

I worked as a data science intern with i3 infosoft pvt ltd for 3 months.

I am a Btech graduate in computer science and engineering from J.C Bose university of science and technology . I Completed my high school from D.A.V public school, passing with an A grade and was awarded the school color for being a high achiever for 5 consecutive years.

I am pursuing a professional data analytics certificate from google , ML from Stanford university through an online course on coursera. I have already completed some other courses such as AI for everyone from the same platform .

I am excited to learn about new technologies and possibilities of the application of AI and machine learning .Data analytics and data science are the fields that have my complete attention at the moment .

Table of Contents

Udemy Project Description

● Situation:

This project is done for the education tech company Udemy to analyze the data on courses from different topics to understand where opportunities to increase revenue may lie, and help the company to track the performance of courses.

● Task:

you have been provided with data on courses from different topics to understand where opportunities to increase revenue may lie, and track the performance of courses.

● Action:
data cleaning
Data processing
Analysis
Finding Insights
Presenting the insights with solution

● Result:
With these data driven insights we have clear evidence that paid courses in web development are the most demanded courses that are capable of generating high revenue for the company. Therefore, I suggest if we increase the cost for Web Development Course it can help the company in increasing their next quarterly earnings.

The Problem

You're a Data Analyst working for the education tech company Udemy. You have been asked by your manager, Head of Curriculum at Udemy, to present the data on course revenue, and you have been provided with data on courses from different topics to understand where opportunities to increase revenue may lie, and track the performance of courses.

Your manager has suggested encouraging Web Development courses to charge more because she believes that these are the most popular courses. She needs to send a report to the CEO in the next three weeks on how they will increase their next quarterly earnings.

Questions to ask to analyse the problem better

- what are the top 20 subscribed courses ?
- How the subscribers were distributed in various subjects?
- What was the level of users in different subjects ?
- How many users used free vs paid courses?
- What is the Sum of subscribers in various subjects ?
- What is the average number of subscribers in various courses?

Design

The data used for the analysis contains 3672 distinct records for 16 different parameters which includes

- course_id
- course_title
- Url
- Free/Paid
- Free Beginner Course
- Price
- num_subscribers
- num_reviews
- num_lectures
- Level
- Rating
- content_duration
- published_timestamp
- Subject
- Date

The data was cleaned in excel then, pivot tables were made to understand various aspects of it .

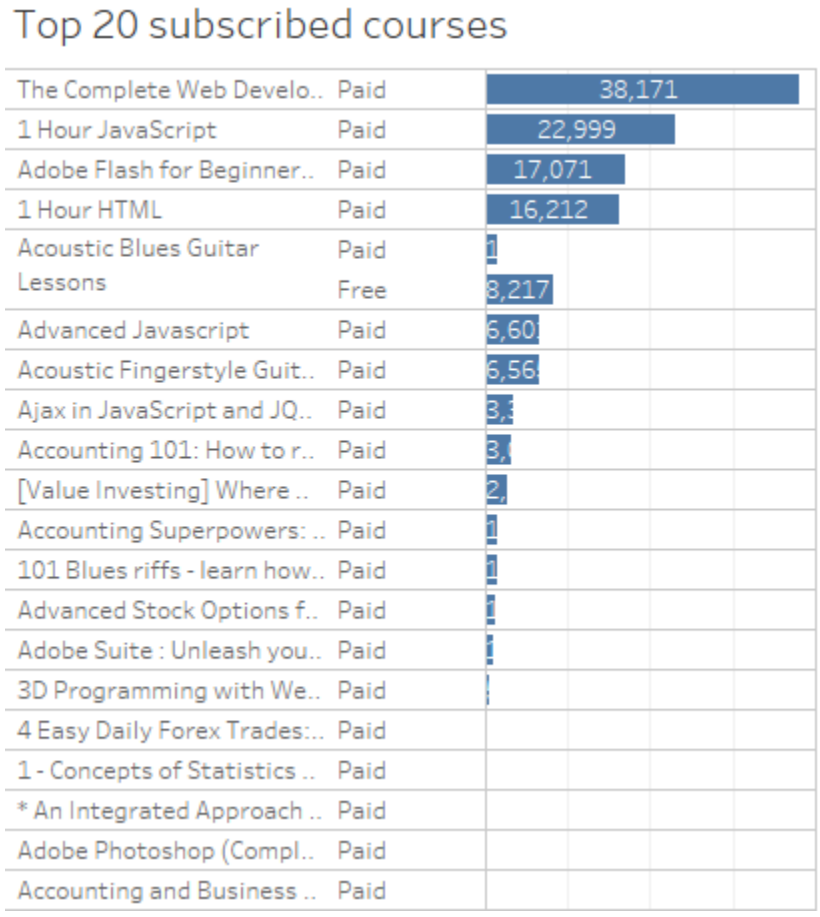
The cleaned data was then imported on tableau to relate the different columns of data to get answers to our desired questions in a visualized manner .

Findings

This section contains answers to the questions that were raised in the previous section.

Finding 1

Figure 1: Total Subscriptions by Category



This bar graph represents the top 20 subscribed courses by users on Udemy, complete web development course being on the top and Accounting and Business scoring the 20th position.

It also shows a very interesting pattern that most of the top 20 subscribed courses were actually the paid courses. Which further infers that user are willing to pay the enrollment fee if the course fulfills their requirement.

Figure 2: paid vs free courses

As we can see in the pie chart above 90% courses available on Udemy are paid courses.

Table 1: Free/Paid Users

Free/Paid
Free 310
Paid 3362

Finding 2

Table 2: Sum Subscribers by Subject

subject

Business Finance 1868711

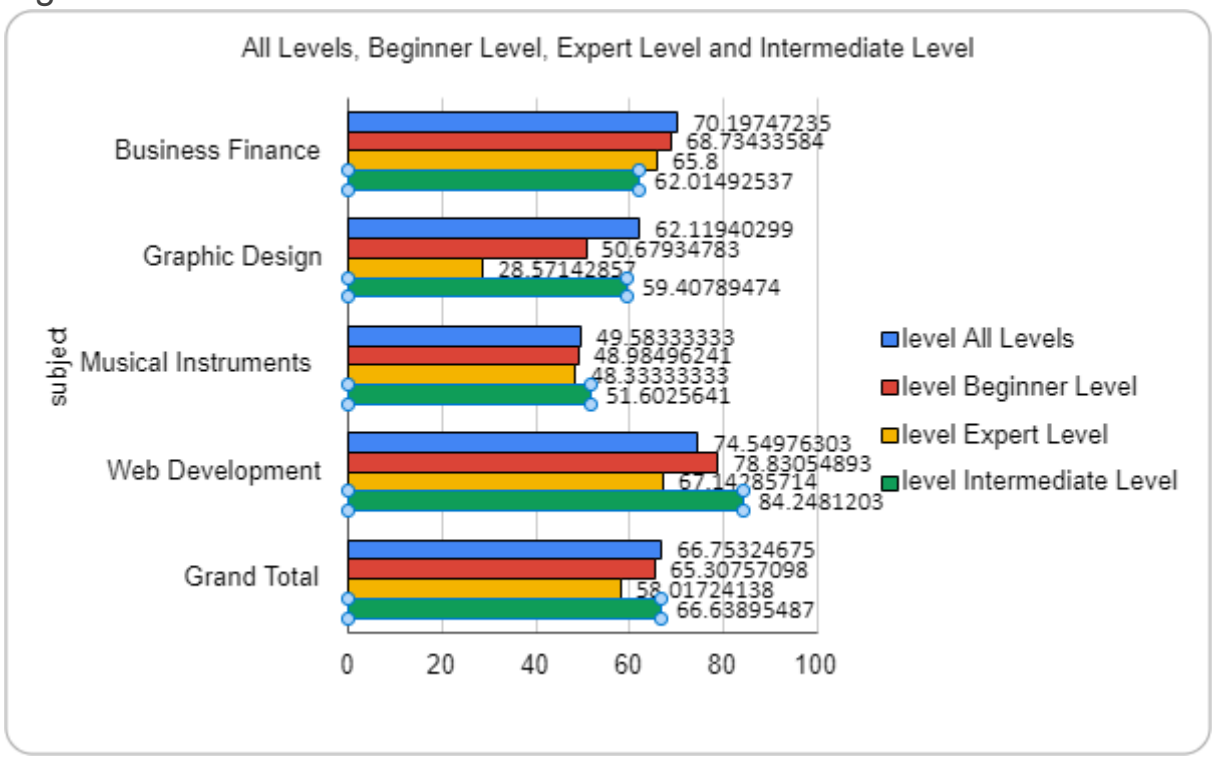
Graphic Design 1063148

Musical Instruments 846689

Web Development 7937287

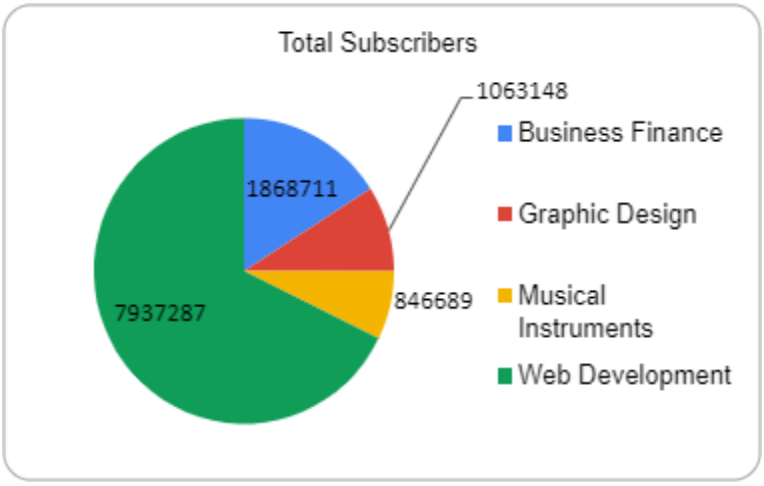
Grand Total 11715835

Figure 3: level wise distribution



The bar chart shows that very few users are enrolled in expert level skills and most of them are enrolled in all levels.

Figure 4 : subject wise distribution of subscribers



The pie chart depicts that the most preferred choice of subject among the users is web development.

Finding 3

Figure 4: 10 top rated courses

These are the top ten Highly rated courses by the users.

Conclusions

With these data driven insights we have clear evidence that paid courses in web development are the most demanded courses that are capable of generating high revenue for the company. Therefore, I suggest if we increase the cost for Web Development Course it can help the company in increasing their next quarterly earnings.

Bank customer churning Project Description

Dataset source : https://www.kaggle.com/sakshigoyal7/credit-card-customers

The problem

Background:

A manager at a local bank is disturbed with more and more customers leaving their credit card services. They need a way of predicting which customers are most likely to stop using their credit card products (Customer Churn) in order to proactively check in on the customer to provide them better services in order to convince them to change their minds. You are given a dataset of 10,000 customers with 18 features per customer. Roughly 16% of the current customer base have churned so far, so it will be difficult to predict the ones who will.

As you analyze the data, before you create the model, the sales team also needs you to determine the most influential factors that can lead to a customer's decision of leaving the business. The head of the sales department is expecting a report that helps them visualize where the differences lie between churning and non-churning customers.

OBJECTIVE:

- Identify which customers are most likely to be churned so the bank manager knows who to provide a better service to increase the chance of retaining the customer
- A clean and easy to understand visual report that helps the sales team better visualize what makes a client churn or not churn

Data Design

Cleaning:

• The dataset contained some 2unwanted columns which were deleted before preceding further in the analysis phase.
1.Naive_Bayes_Classifier_Attrition_Flag_Card_Category_Contacts_Count_12_mon_Dependent_count_Educat ion_Level_Months_Inactive_12_mon_1
2.Naive_Bayes_Classifier_Attrition_Flag_Card_Category_Contacts_Count_12_mon_Dependent_count_Educat ion_Level_Months_Inactive_12_mon_2
• Column Avg_Utilization_Ratio contained an entry error ie 10.12% was recorded as 0.1012

The data was cleaned in excel then, pivot tables were made to understand various aspects of it .

The cleaned data was then imported on tableau to relate the different columns of data to get answers to our desired questions in a visualized manner

Findings

Figure 1: ATTRITION FLAG BASED ON CUSTOMER

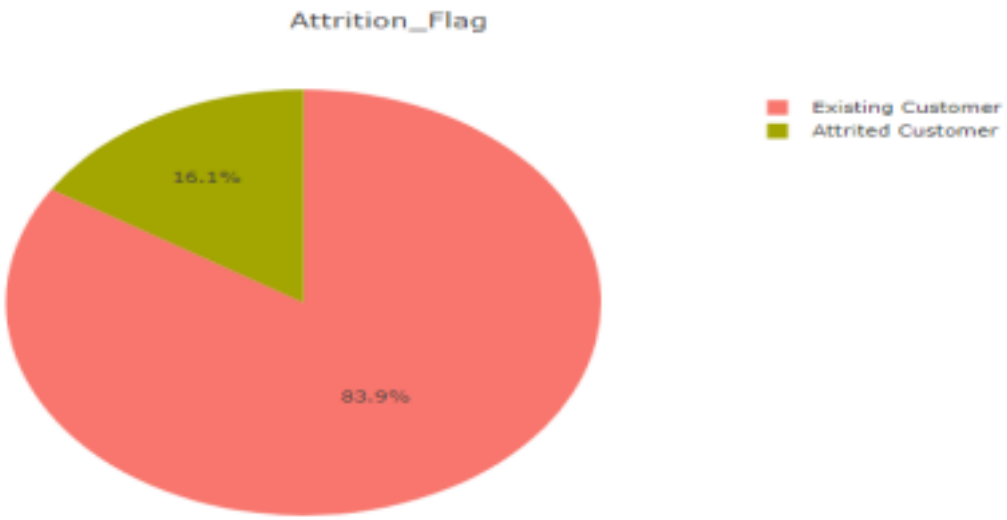


Figure 2: BASED ON GENDER

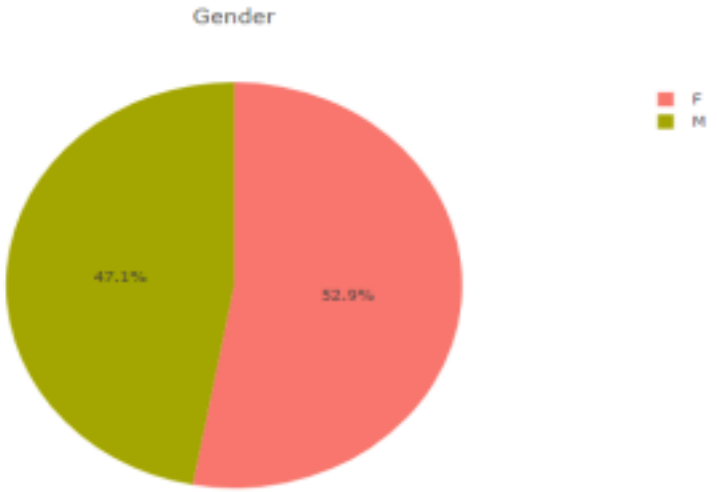


TABLE 1:

Card_Category	SUM of Total_Relationship_Count
Blue	36 31 6
Gold	349
Platinum	46
Silver	1899

Figure 3: CARD TYPES WISE DISTRIBUTION

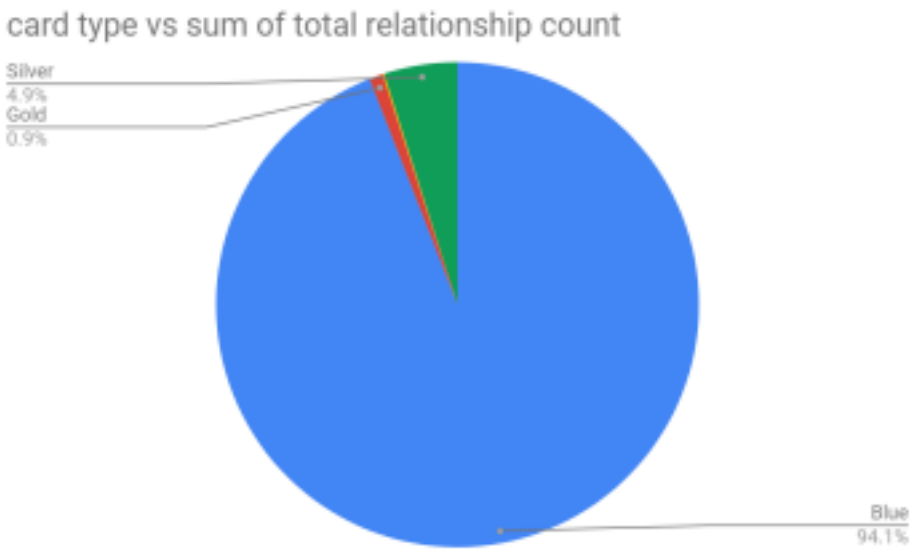


TABLE 2:

Income_Category	SUM of Credit_Limit
$120K +	14334493.5
$40K - $60K	9777445.1
$60K - $80K	15083799.7
$80K - $120K	24268182.6
Less than $40K	13369434.5
Unknown	10582439.7

Figure 4:

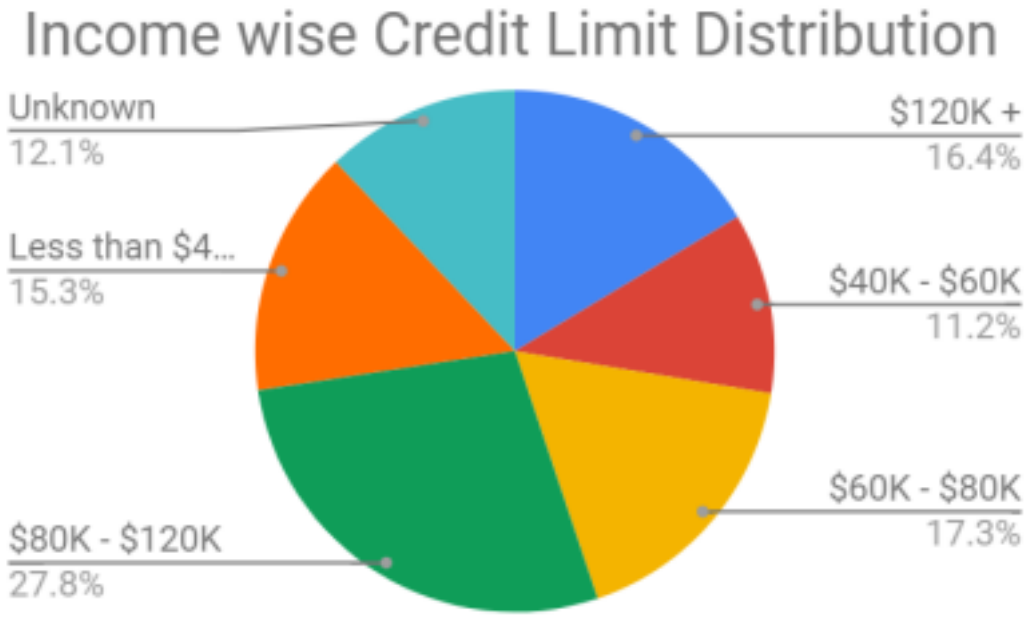


Figure 4:

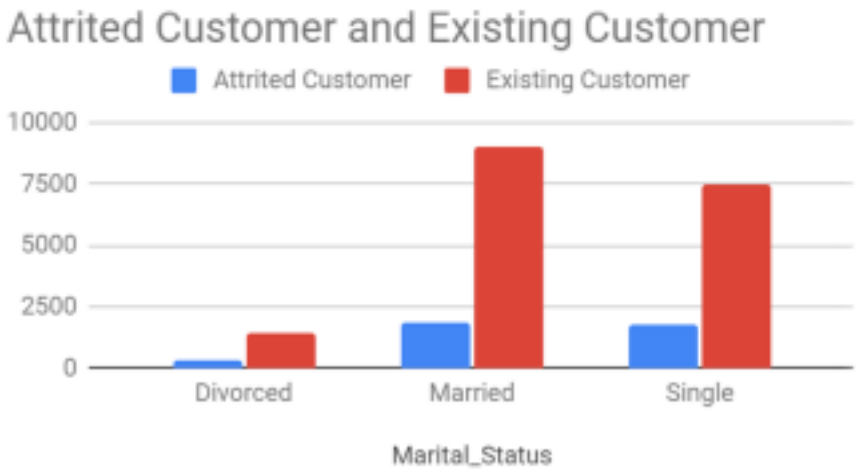


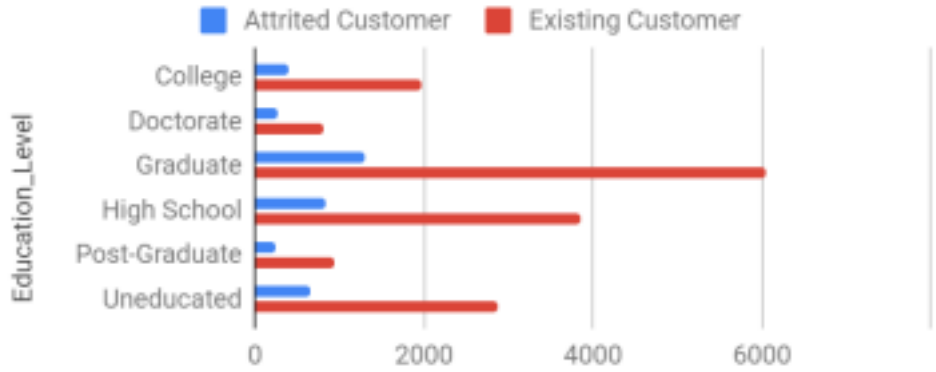
TABLE 3:

SUM of Months_Inactive_12_mon	*Attrition_Flag*	

Education_Level	Attrited Customer	Existing Customer
College	405	1979
Doctorate	257	804
Graduate	1312	6039

High School	832	3858
Post-Graduate	248	946
Uneducated	653	2865

Figure 5:



The customer gender is almost even, 30% college graduates with half being either Highschool graduates, unknown, or uneducated. The remaining 40% are either current college students, or grad students.

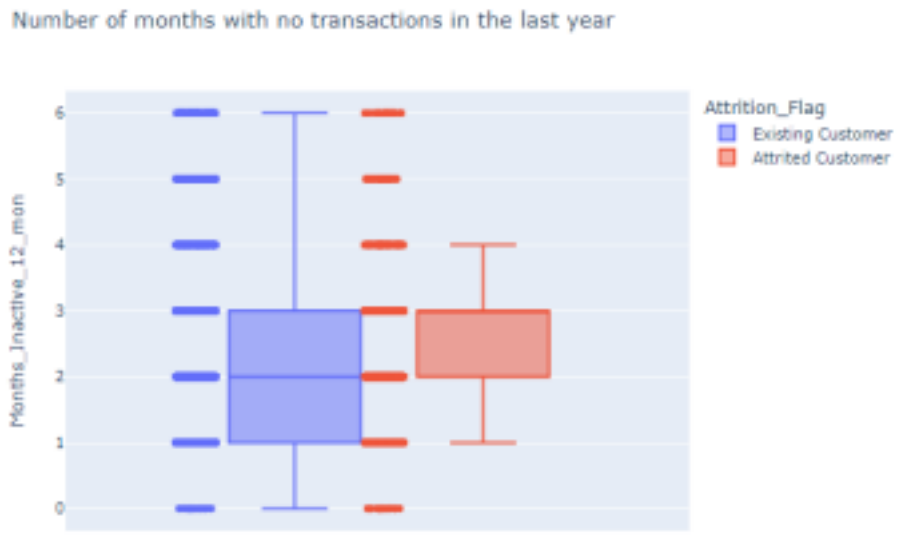
Almost half are married, 38% single, and the remaining 12 are divorced or unknown. 35% of customers make less than $40k per year which is near the poverty threshold. The rest are more evenly spaced out. 93% of customers choose the cheapest card option (likely the lowest interest rate) with a tiny portion choosing the more expensive cards

Data Analysis

Figure 6:

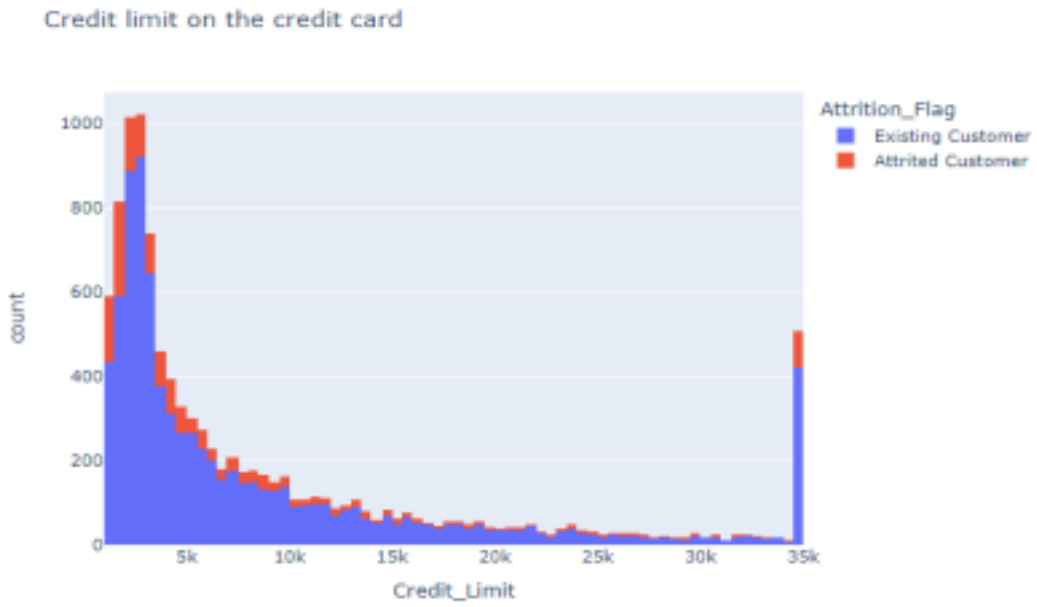
Churned customers are likely to hold less credit cards than existing customers which is shown by a lower median. Is there a deal you provide that favors customers with multiple credit cards? (Like customers with spouses, families, or Business that need additional cards)

Figure 7:



Churned customers tend to have slightly more inactive months, but the distribution is more concentrated from the 1-4 months inactive.

Figure 8:

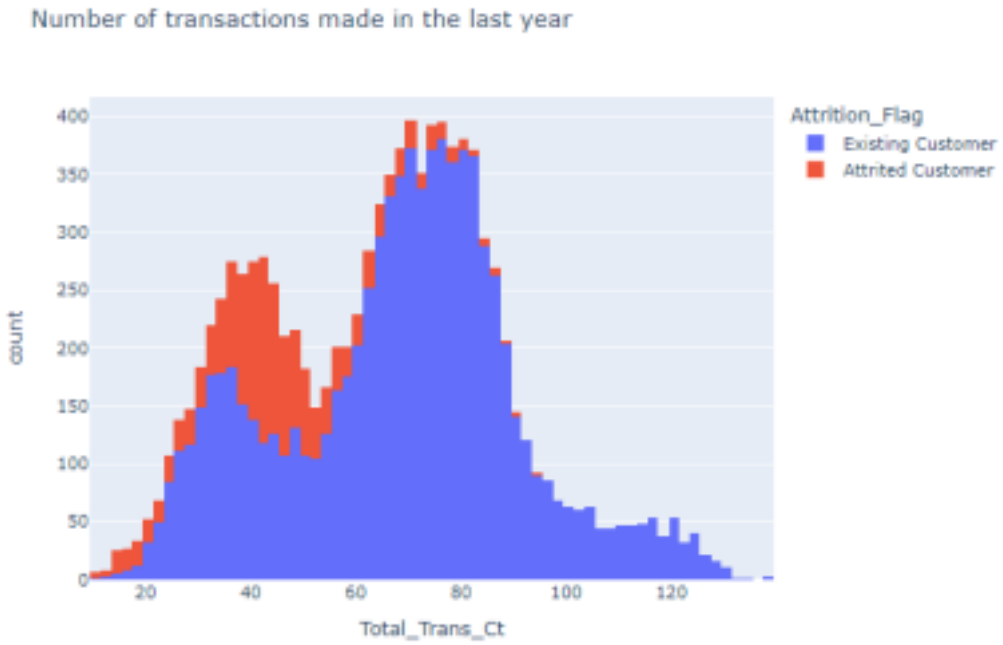


Churned customers have a lower credit limit, so perhaps increase the credit limit for them.

Figure 9:

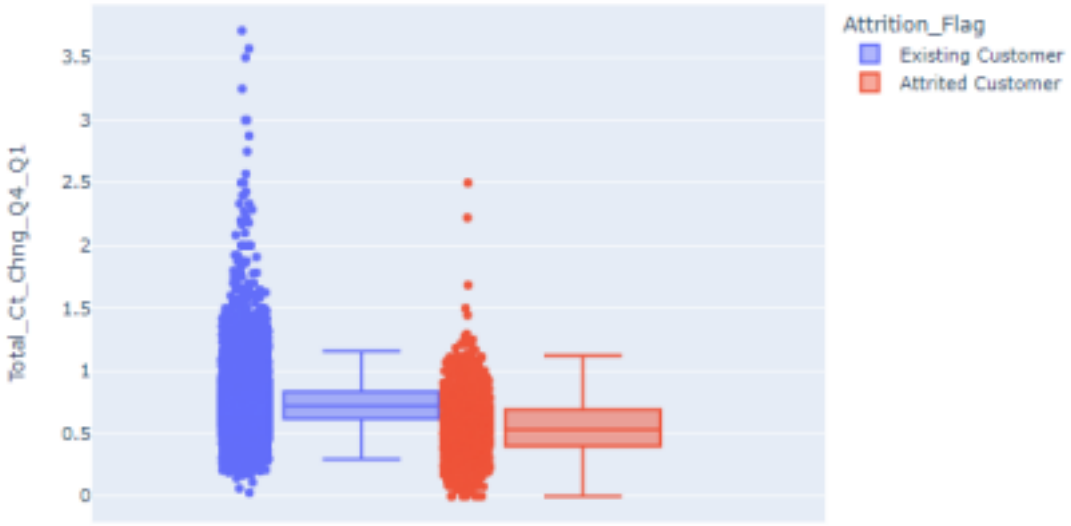
Churned customers have a much smaller revolving balance which, because they don't fully pay off their credit card balance, may signify that they have less disposable income than staying customers that know they can pay off their revolving balance.

Figure 10:



Churned customers will have a lower number of transactions, which makes sense as they're less involved with this company and will have a smaller transaction change over time as displayed below.

Figure 11:



Conclusions

Total Transaction change, revolving balance, and Number of contacts within the past year are most correlated with a churning customer.

To improve the customer retention-rate the company should work towards keeping a check on factors causing the above factors and should come up with innovative solutions to improving their services for these customers.