

DATA ANALYSIS PORTFOLIO

ABOUT

It has been rightly said by Marcus Aurelius that “nothing has such power to broaden the mind as the ability to investigate systematically and truly all that comes under the observation in life”.

Hello, I am Diana Satapathy, working as an Assistant System Engineer-Trainee for Tata Consultancy Services from November 2020 – Present, is a process-oriented individual with a keen interest in the field of data analytics. Graduated from Koustuv Institute of Self Domain in 2020 with a Bachelor’s Degree in Computer Science and Engineering.

A competent SQL coder with hands-on experience in python, Tableau, Power BI, Microsoft Excel. Always driven by curiosity and seeks to strike a balance between analytics, problem-solving, business, and creativity. Proficient in interpreting and analyzing data to drive growth. I as an individual have been truly benefitted from observing happenings around me and analyzing them from a fresh perspective that polished my analytical skills. Skilled in collecting, organizing, interpreting, and disseminating various types of statistical figures, I always enjoy working with data.

TABLE OF CONTENT

1. Professional Background

2. Table Of Content

3. Udemy Project Description

The Business Problem

- What Data to be collected?
- What questions to be asked?

Data Design

- Data Cleaning
- Tools Used

Findings

Analysis

Conclusion

4. Financial Complaints Analysis Project Description

The Business Problem

- What Data to be collected?
- What questions to be asked?

Data Design

- Data Cleaning
- Tools Used

Findings

Analysis

Conclusion

5. Appendix

UDEMY PROJECT DESCRIPTION:

SITUATION: As a Data Analyst working for the education tech company Udemy, I have been asked by my manager, Head of Curriculum at Udemy, to present the data on course revenue, and was provided with data on courses from different topics to understand where opportunities to increase revenue may lie, and track the performance of courses.

TASK: Important task of this project as stated by the stakeholder was to come up with strategies and ideas to increase the quarterly revenue and understand what category of courses were the most purchased by people and what other actions could be taken to attract more people to subscribe for the courses provided.

ACTION: As part of the analysis, the data was collected on different subjects which contained their ratings, content duration, and number of subscribers. The data was cleaned in order to avoid any misleading outcome. Then the clean data was plotted in tableau for generating charts to share with the stakeholders such that a better data driven approach could be obtained.

RESULT: The results obtained were as such that more people were subscribing for web development paid courses. Even though Graphics Designing had a healthy subscriber count and good rating the revenue generated was less. The same goes for business courses as well. As a suggestion to the analysis, we could come up with a recommendation system where people picking up certain courses would be presented with a set of similar courses as a recommendation on their feed and this would attract their attention to subscribe for more such courses. Online advertisements and campaigns for bonus courses and discounts can attract new audience as well. To further the analysis, a more detailed report could be generated which would include the age group of the subscribers which would be beneficial for the company to target the specific audience.

THE BUSINESS PROBLEM

As a Data Analyst working for the education tech company UDEMY, I have been asked by my manager, Head of Curriculum at Udemy, to present the data on course revenue. I have been provided with data on courses from different topics to understand where opportunities to increase revenue may lie, also track the performance of courses.

The business problem as stated by the stakeholder is how can we increase their quarterly sales and what category of courses should be promoted in order to boost their earning. The manager has also added that we should charge more for the web development courses because according to her these were more popular among learners than the others.

WHAT DATA SHOULD BE COLLECTED TO UNDERSTAND THIS PROBLEM STATEMENT?

- Data related to every category of courses
- Number of subscribers for each subject.
- Number of subscribers for each course.
- Number of subscribers for each level of courses (All/Beginner/Intermediate/Advanced)
- Number of free/paid courses by subjects.
- Number of free beginner courses by subjects.
- Content duration of each course
- Content duration of each subject category.
- Number of reviews on each course and also on each category of subjects.

WHAT QUESTIONS TO ASK?

- What would be the targeted quarterly revenues?
- What type of courses are people most likely to enrol?
- What course category has the greatest number of subscribers?
- Which course charges the most and how many subscribers does it have?

DATA DESIGN

Data comes from multiple sources, and data analysts often gather data from multiple sources and combine these for data analysis. The process of gathering data from these sources and presenting it is called data consolidation. Data consolidation is a crucial step, as the accuracy of the insights from your data analysis depends heavily on the quality of data used. Here we have been provided data of all the udemy courses by subjects which includes course price, number of subscribers, duration, published date.

DATA CLEANING

Clean data is essential when using data to derive insights and make business decisions.

Garbage in, garbage out, is a concept common to computer science and mathematics that can be applied in Data Analysis - the quality of output (your insights generated) is determined by the quality of the input.

This is where Data cleaning comes in - Data cleaning is the process of preparing data for analysis by removing or modifying data that is incorrect, incomplete, irrelevant, duplicated, or improperly formatted. Data cleaning plays an important role in the analytical process and making sure that the answers we uncover are reliable and of high quality.

Below mentioned functions were used to clean data.

1) Remove duplicates

Select the entire data sheet data to remove duplicates from.

Data > Remove Duplicates

2) Removing blank cells

Select the entire sheet or dataset and go to Data>Create a filter

Click on the Filter icon at the top of any column, then click on Filter by condition and select 'is empty'.

Blank cells will arise to the top of the sheet and can be removed.

3) Headers

Ensure you have clear and concise names for headers and use dashes or underscores in between words to make it easier to parse later on.

4) Find and replace

If you examine the data, you will see that the Web Development subject title is not the same as other subject titles. Use the Find and Replace function to make the Web Development subject consistent with other subjects.

Edit > Find and Replace

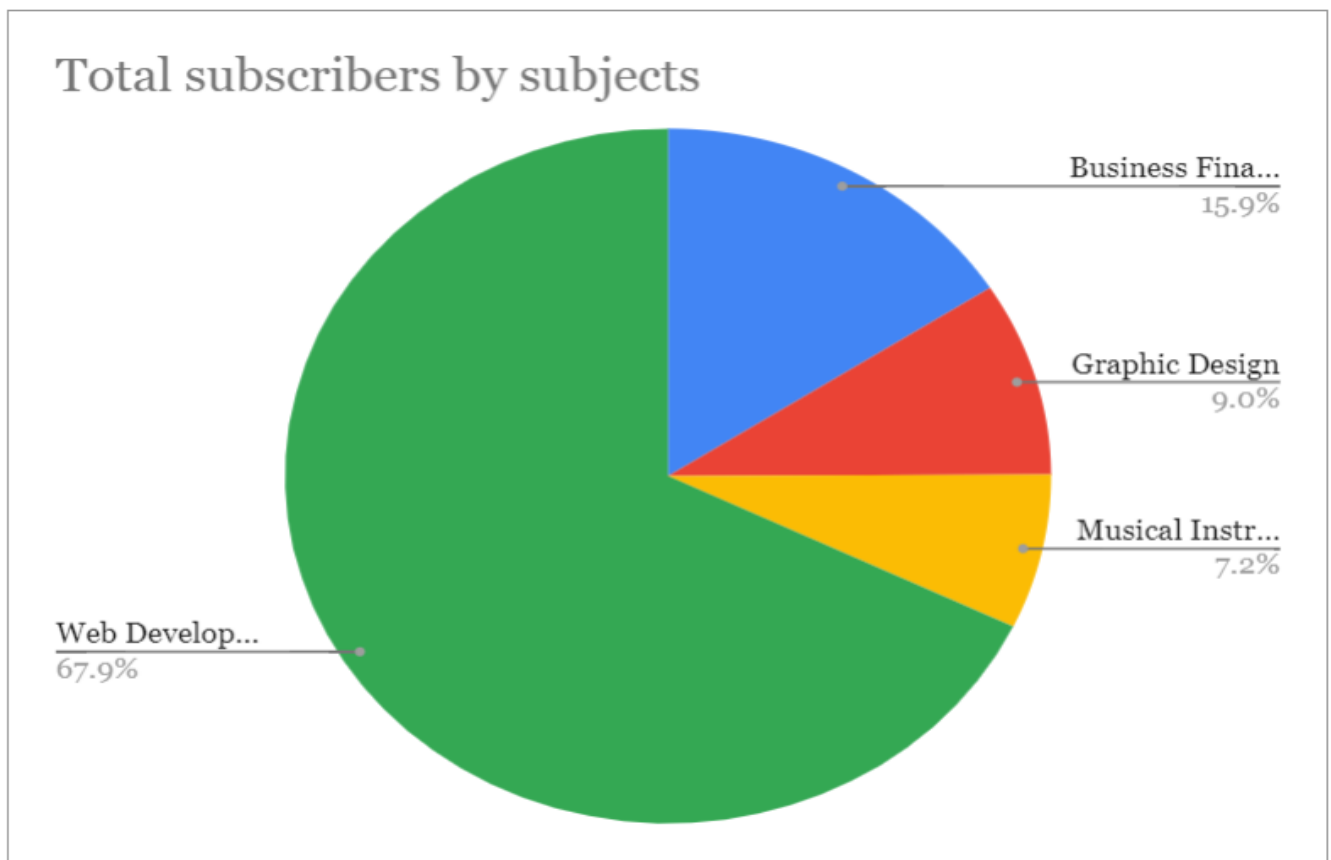
TOOLS USED

Google Sheets: For initial data cleaning and basic analysis.

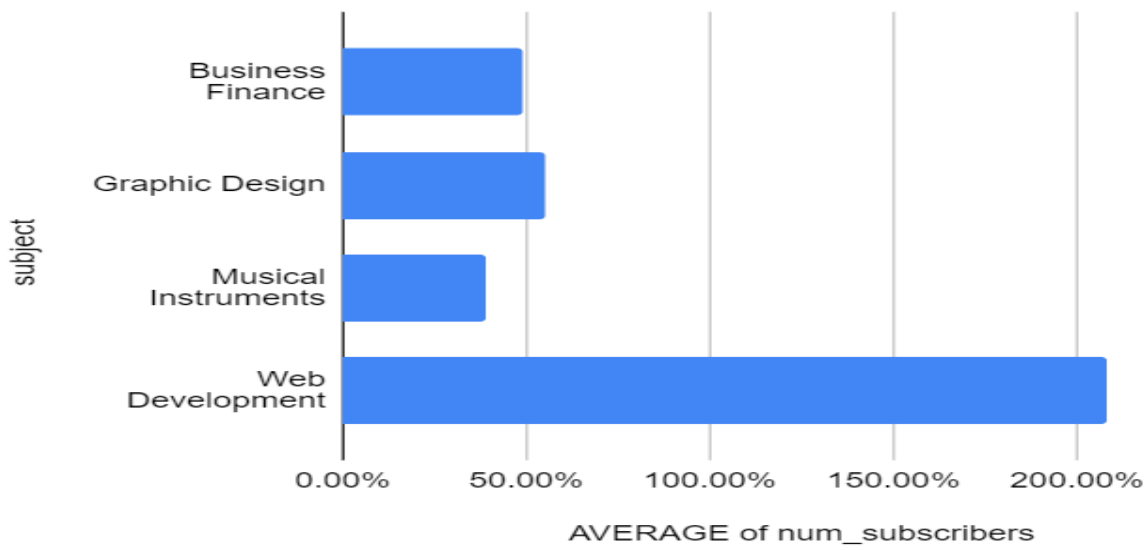
Tableau Desktop: For visualization of datasets.

FINDINGS

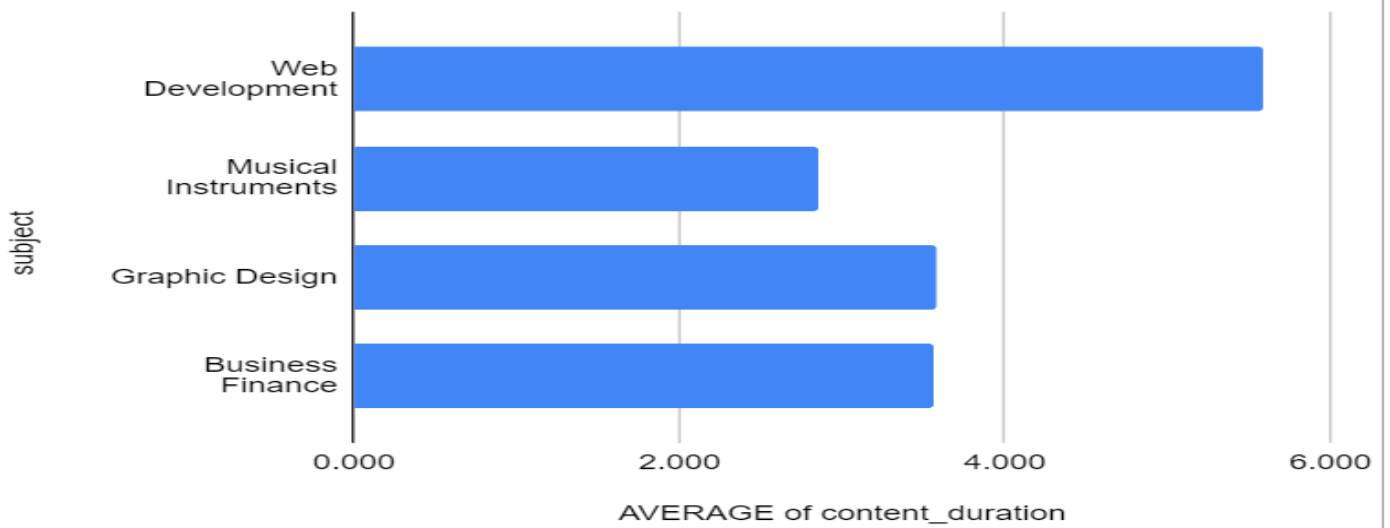
- As part of the analysis, we first created a tree map representing the hierarchy of all the four subjects by number of subscribers based on the price value of the courses whether free or paid and we find that there were more subscribers for the web development paid courses and least for musical instrument courses.
- From our insights also we can find that more and more people are subscribing for the Web development courses, followed by Graphics Design and Musical Instruments Category being the least subscribed too.
- We skimmed down to the top 20 courses by subscribers by use of sets in tableau to understand which of the courses were most purchased by the people and we find that out of the 20 courses, 18 were from web development again, 1 from business finance and 1 from musical instruments. This indicates that many people are inclined to purchasing more development courses and hence there are chances that we could attract more people from tech to sign up for similar courses.



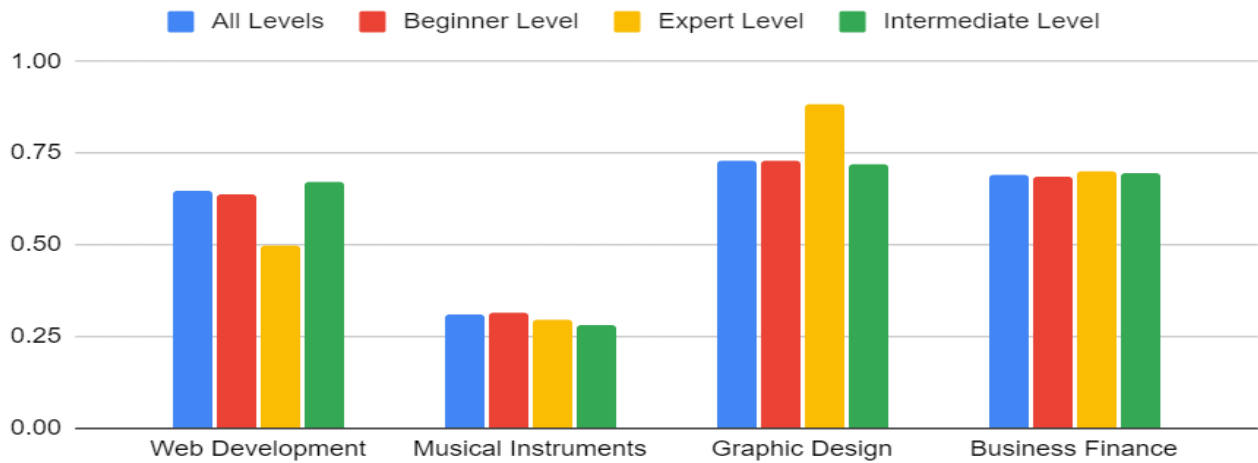
AVERAGE of Subscribers vs Subject



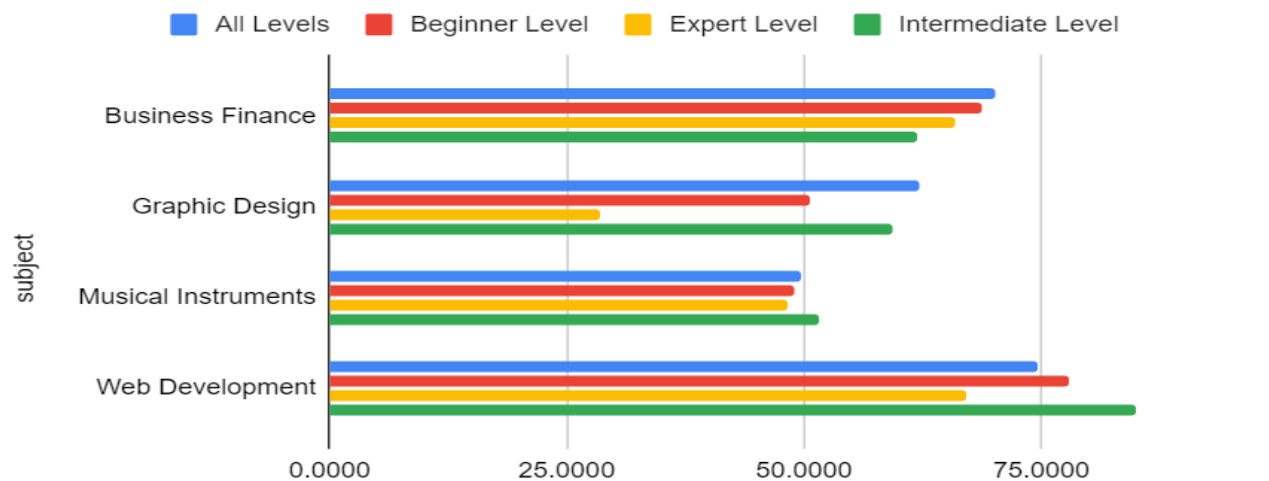
AVERAGE of content_duration vs subject



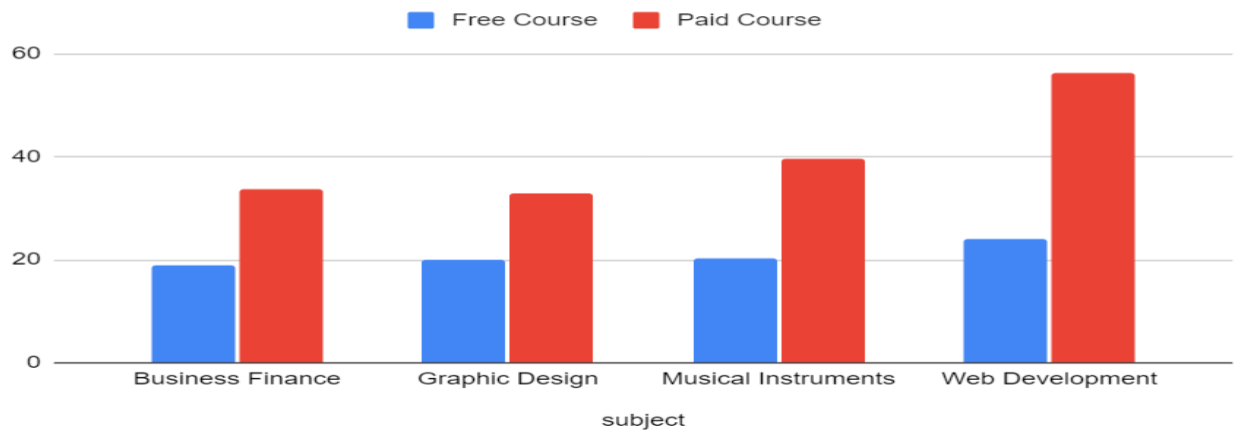
Avg Rating Of Subjects By Levels



AVG COST PER SUBJECT BY LEVELS

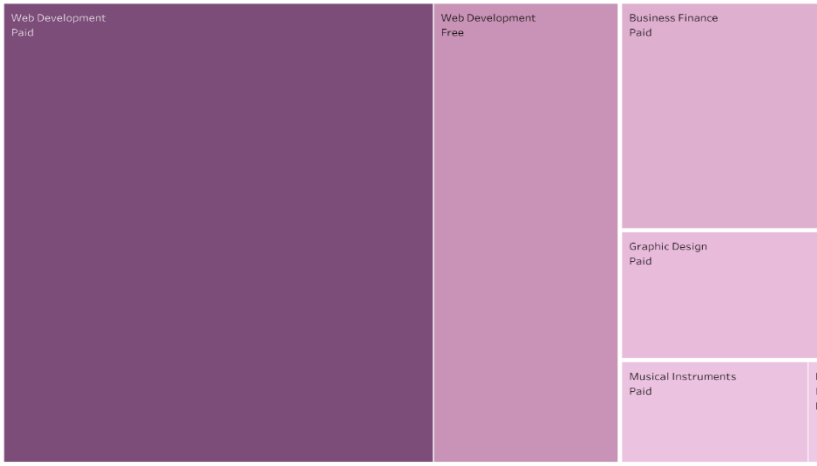


Free Course and Paid Course

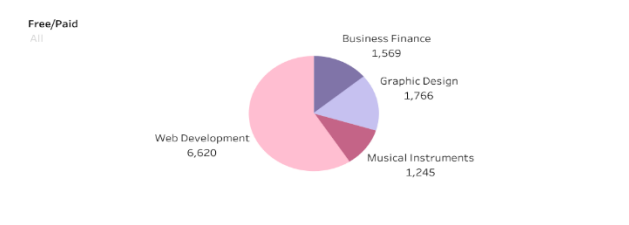


UDEMY COURSES OVERVIEW

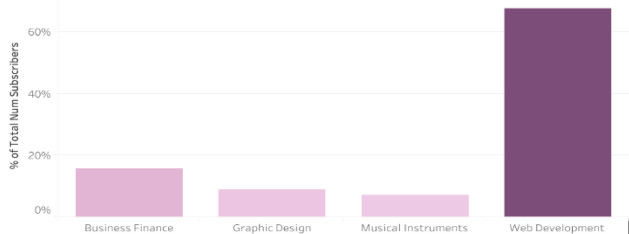
Number of Subscribers by Subjects



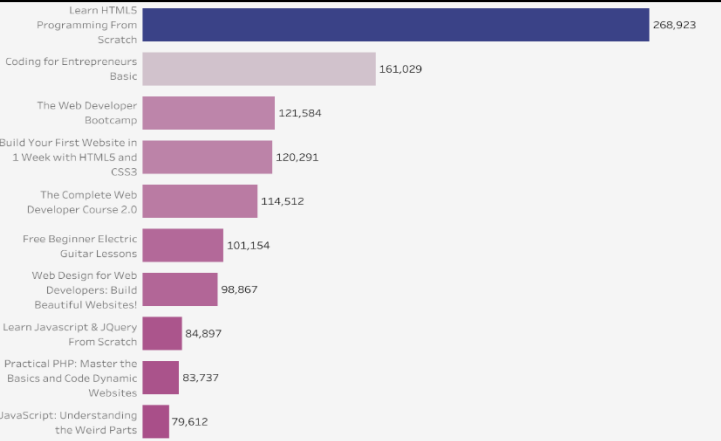
Subjects by AVG Number of Subscribers



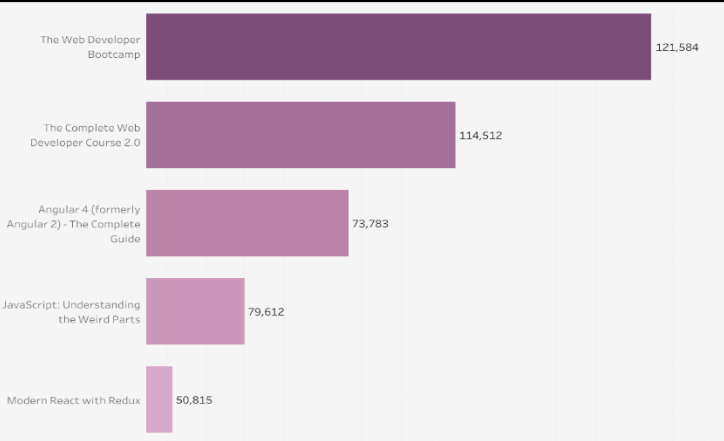
Subjects by percent of total number of subscribers



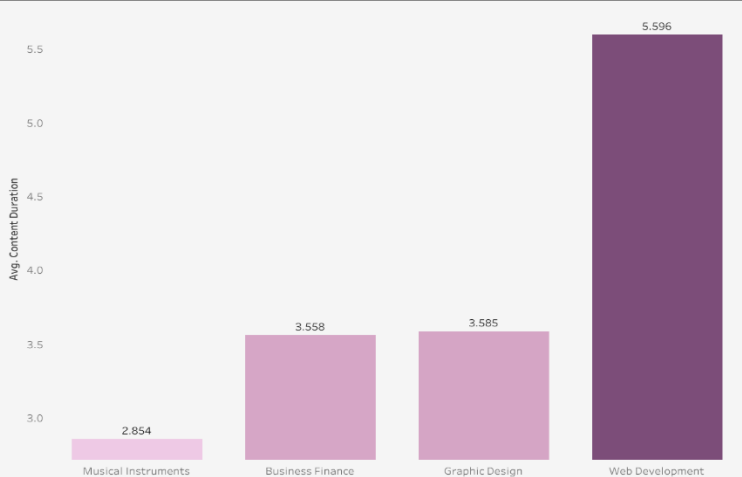
Top Courses by Number of Subscribers



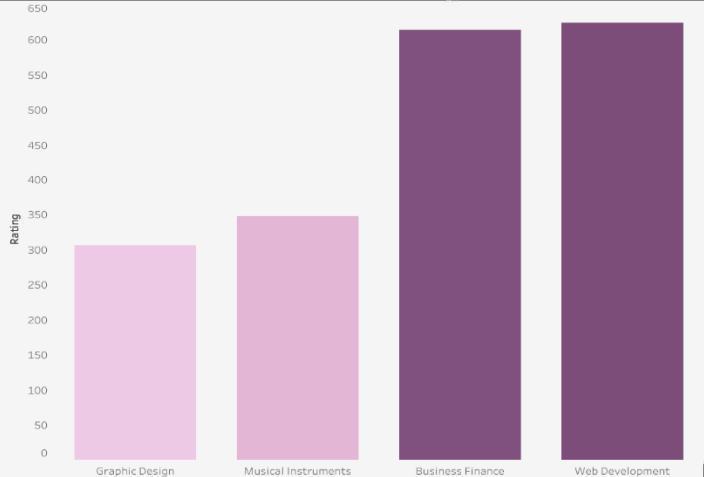
Top Courses By Number of Reviews



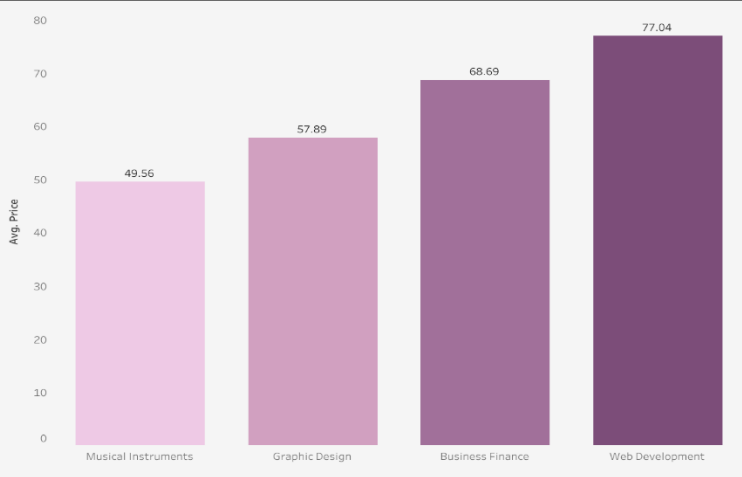
Subjects by AVG Content Duration



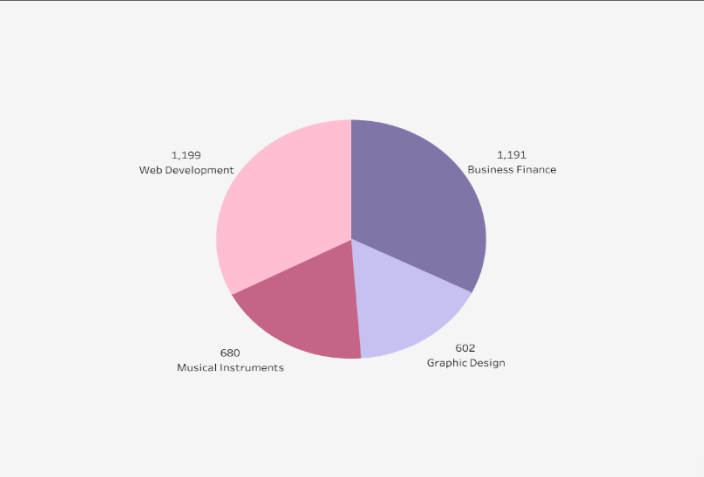
Subjects by Rating



Subjects by AVG Price



Count of Free Beginner Course By Subjects



ANALYSIS

1. Why do stakeholders want to get insights on the revenue?

The main reason behind asking was to gather more information on how the company has been performing in the ed tech world and hence where and all should they work to generate greater profits.

2. Why is web development the most purchased course?

Technology being the highest paying job in the market has attracted people from different background to upskill and generate better employment thus advancing in their career.

3. Why is musical instrument course at the lowest even though it has decent rating?

A fact that music is best learnt in physical environment is also a cause why people aren't signing up for the subject through an online medium.

4. Why is graphical design subject having less subscribers?

Graphics design is another booming career opportunity in the market. An individual with web development knowledge and a bit of designing can actually make great success. But considering that these courses are not much popular amongst the subscribers there could be room for improvement. Also, Average ratings of Graphics Design courses at each level is more than that of all the other three subjects which indicates graphics design courses does have some quality content but there is a need to market in well in order to attract audience.

5. Why are business finances generating less revenue even if it has the second highest number of subscribers?

Again, there is a lack of audience which means we need to adapt better marketing strategies in order to expand our growth.

CONCLUSION

- Adding more courses to the business finances subject with some bonus add-ons might probably attract more subscribers.
- Since Web development courses and business finance are the most subscribed, we can increase the price for the courses from both the categories to generate more revenues. To attract more subscribers, adding bonus mini courses on completion of the courses would be a benefit.
- Practical examples are boon hence it'll add value to the course content if there were more project-based courses in all the four subjects as this would definitely attract more and more subscribers.
- Graphics Designing courses are also having a healthy subscriber and boosting their sales would add more revenue. This can be done by adding discounts and bonus resources.
- We need to understand our subscribers, from where they come, what kind of content they prefer, what age group are they from and what are their goals and using this information, we could provide them with a list of all the relevant courses to purchase. Better marketing strategies, online presence and recommendation system would really add towards the organization's growth and hence generate more profit.

FINANCIAL COMPLAINT PROJECT DESCRIPTION:

SITUATION: CUMULUS FINANCE, a financial consultant firm that provides credit card services, debit card services, loans, account related services, mortgages etc. To understand their consumers and how well they have been functioning with the audience, they wanted to run an analysis on all the financial complaints registered against their services, what issues were the consumers facing, how well has the team been performing in resolving the issues and what were the outcomes i.e., customer satisfaction, whether the consumer was satisfied or did they raise any dispute, was the settlement monetary or non-monetary. Thus, as a part of the customer support analytics team, we were asked to make an analysis that answered some questions raised by the stakeholders.

TASK: Important task of this project as stated by the stakeholder was to analyze how has the team been performing in providing service and was there a need to incorporate more people for more timely response. What percent of the complaints were resolved and how many were under process. What issues were the consumers facing and against which products more complaints were registered.

ACTION: As part of the analysis, complaint data was collected from 19th January, 2021 - 20th October, 2020 which included complaint ids, product against which the issue was registered, issues and sub issues faces by the consumers, company's response to the complaints, consumer satisfaction to the outcome of the issue resolved. The data was cleaned in order to avoid any misleading outcome. Then the clean data was plotted in tableau for generating charts to share with the stakeholders such that a better data driven approach could be obtained.

RESULT: The results obtained were as such that there was a total of 75,513 complaints registered with the firm between 19th January, 2021 - 20th October, 2020 out of which 0.38% were under process and 0.62% were resolved. 98% Percent of the issues were responded quickly by the executives which implies that there was barely any delay in addressing the customer's issue. Hence there is barely any need to add more people to the team. 84.4% of the total complaints were resolved without monetary values and only 9.7% had issues with the outcomes rest 90.3% were satisfied or did not provide any review.

BUSINESS PROBLEM

CUMULUS FINANCE, is a financial consultant firm that provides credit card services, debit cards, loans, bank account related services in Kimberley, Northern Cape Province, South Africa. The finance team has encountered that more and more customer has been facing issues with a lot of their services and have received many complaints as well.

They wanted to make sure whether the firm has been able to provide timely service to the customer without any dispute or not.

For this the analytics team was asked to present a report and were provided with customer complaint data which contained complaint entries against products and sub products along with the issues mentioned.

As part of the analytics team, it was our responsibility to come up with a dashboard for providing better insights.

QUESTIONS TO BE ANSWERED

1. WHAT WAS THE TOTAL NUMBER OF COMPLAINTS REGISTERED BY CUSTOMERS?
2. WHICH PRODUCT HAD THE HIGHEST NUMBER OF COMPLAINTS REGISTERED AGAINST IT?
3. MENTION THE ISSUE WHICH WAS FACED BY MAXIMUM NUMBER OF CUSTOMERS.
4. WHAT WAS THE COMPANY'S RESPONSE TO THE CONSUMER?
5. WERE ANY OF THE CUSTOMER DISSAPOINTED BY THE SERVICE AND RAISED A DISPUTE?
6. HOW WERE THE COMPLAINTS MADE?

DATA DESIGN

Data collected initially had a lot of missing values, cleaning null values for which I replaced the blanks to N/A because most of the complaints raised had no feedback from the consumer after the issue was resolved and it wouldn't have given a better result if we'd have removed the null values. Changed the data types of the date when the complaints were registered to Date datatype. Ensured to have clear and concise headers.

TOOLS USED

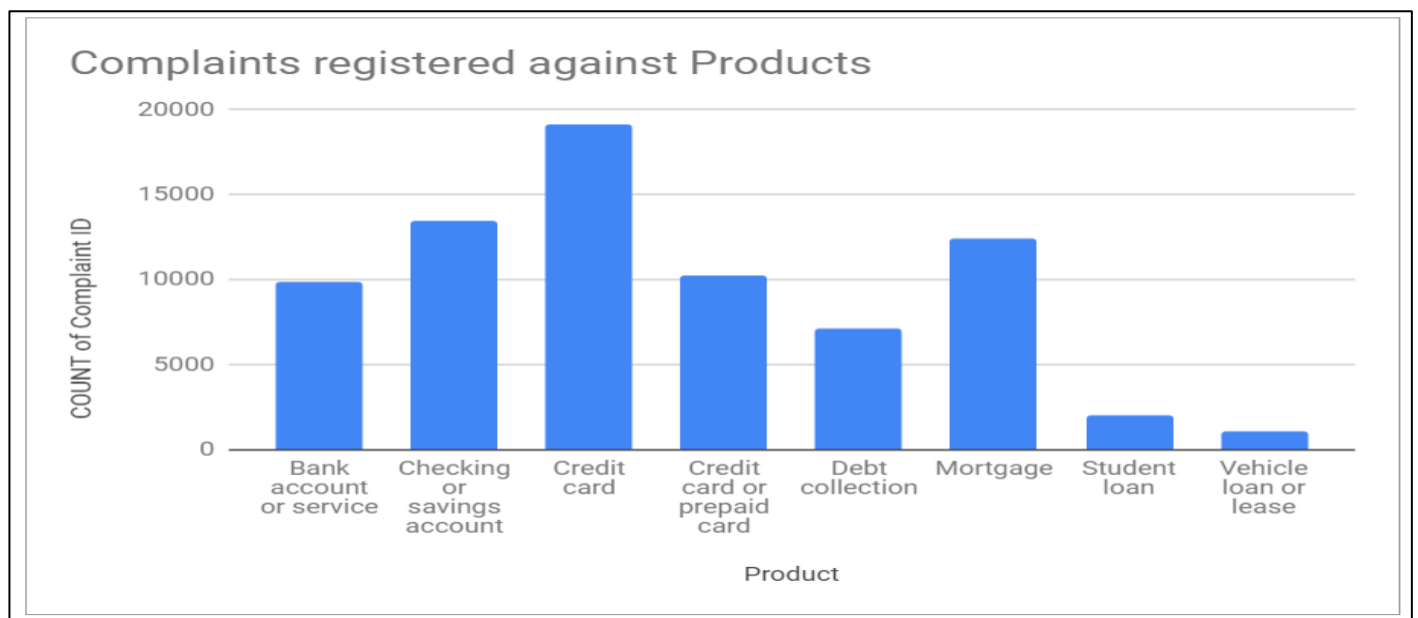
Google Sheets: For initial data cleaning and basic analysis.

Tableau Desktop: For visualization of datasets.

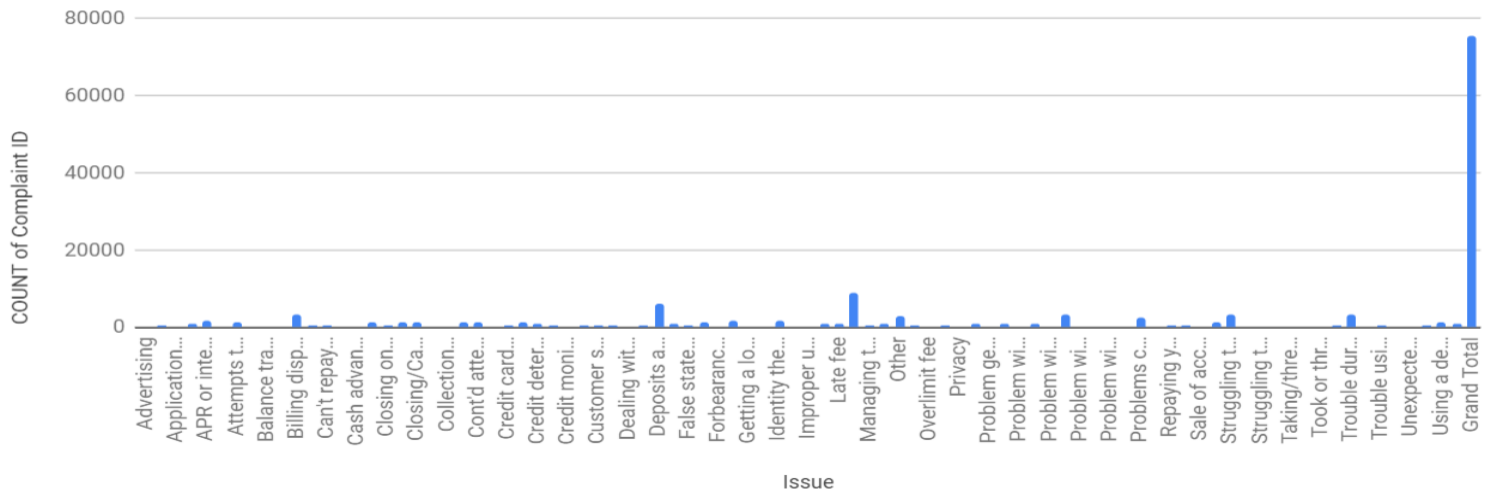
FINDINGS

As part of the analysis, we have gathered some details from the data. Let me walk you through them.

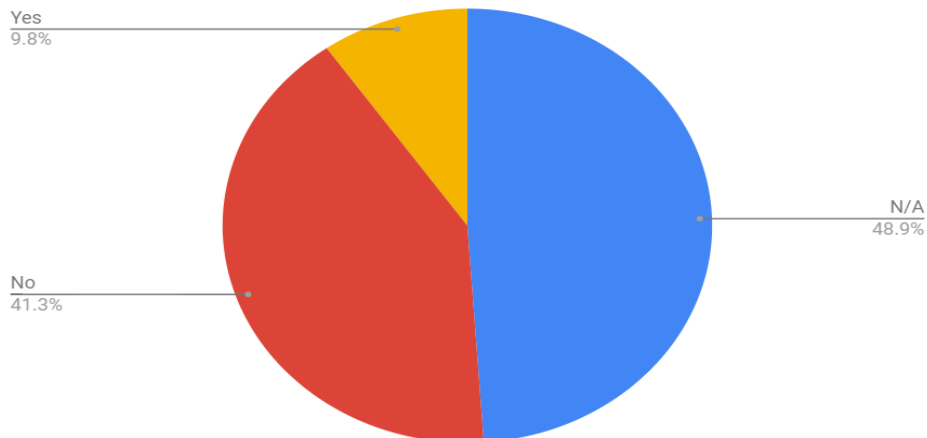
1. Created a KPI which shows the number of complaints filed by the customers between 1/19/2021 – 10/13/2020. There was a total of 75,513 complaints registered with the firm.
2. 98% Percent of the issues were responded quickly by the executives which implies that there was barely any delay in addressing the customer's issue.
3. 0.38% of the total complaints was still under progress and remaining 0.62% was resolved or closed.
4. 9.75% of the customer had raised dispute which accounts for 7363 disputes and is fairly a small amount as the maximum percentage nearly 90% of the customer were satisfied with the outcome.
5. 84.4% of the total complaints were resolved without monetary values.
6. If we were to consider the complaints raised by products, we find that the top three products which had complaints registered against them were Credit card, checking or savings account and Mortgage.
7. The bar chart with complaints by issues indicates that customer faced problems with account management, deposits and withdrawals, trouble during payment process, struggling while paying for mortgage etc.
8. If we were to skim down to issues filtered by products.
 - Issues registered with credit cards were mostly for billing, fraud, identity theft etc.
 - Max issues registered with savings account were account opening, closing and managing of the same.
 - With mortgage there were trouble during payment processing and settlement process.
9. 50% of the complaints were registered via web i.e., from the website, 295 by referrals and the remaining where via posts, fax and emails.



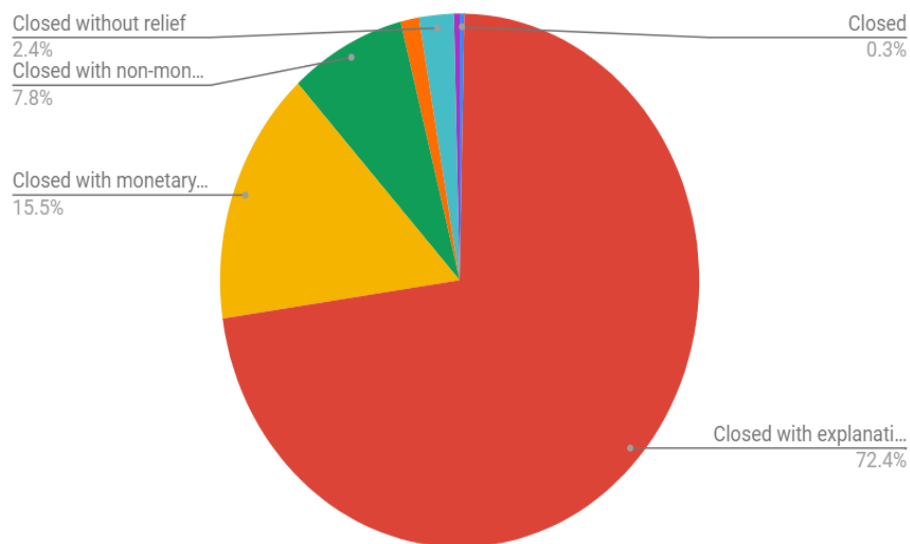
Complaints By Issue



Consumer Disputed

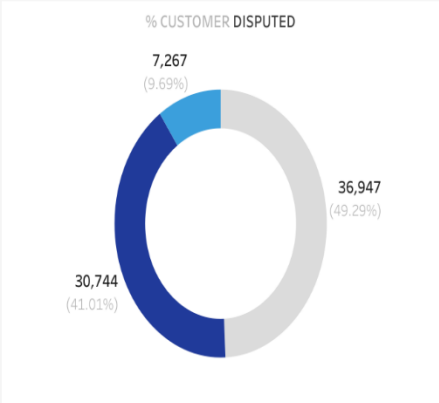
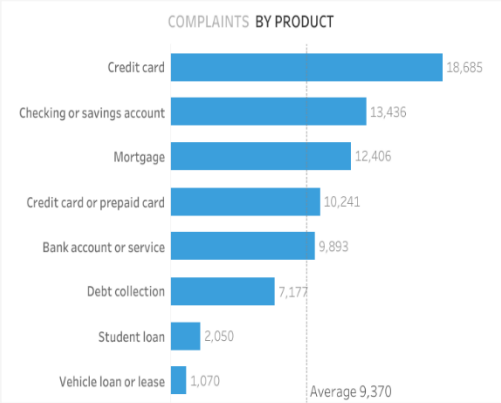
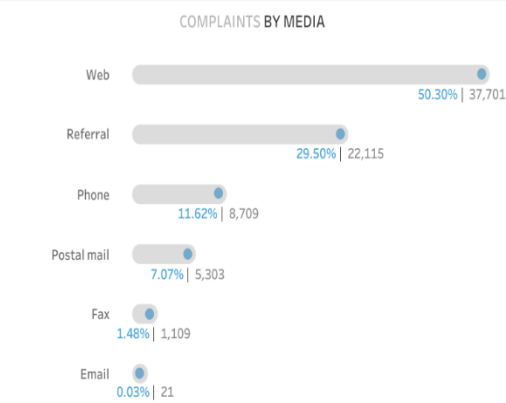
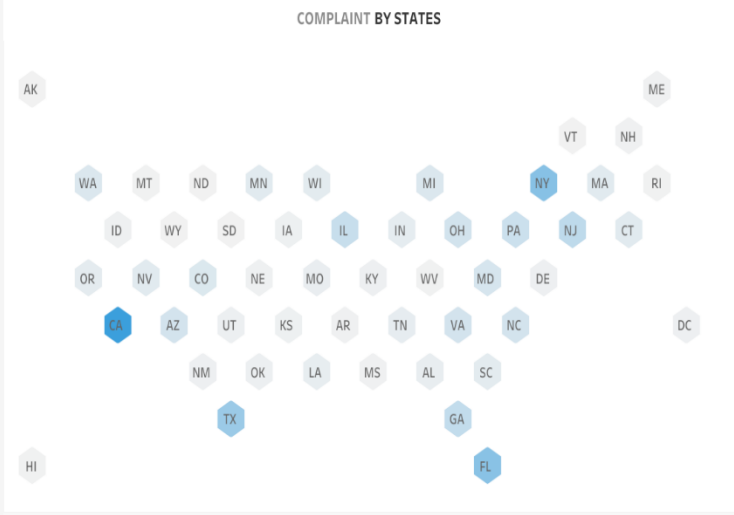
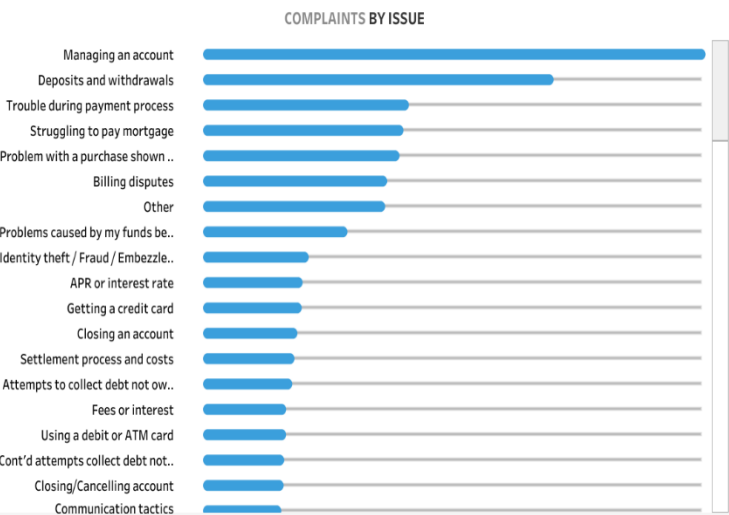
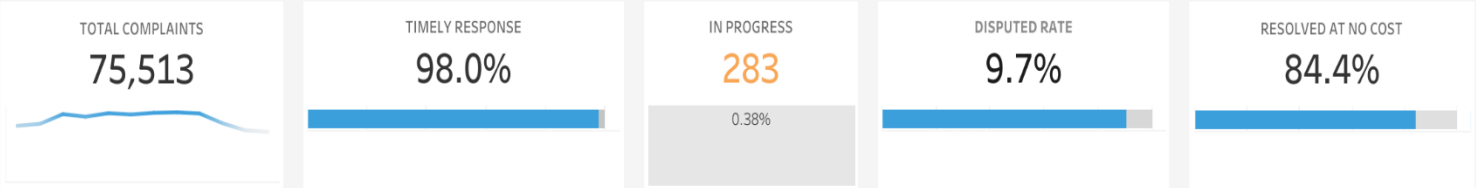


Issues Resolved By



FINANCIAL COMPLAINTS | OVERVIEW

Date Submitted
1/19/2012 to 10/13/2020
and Null values



ANALYSIS

1. Why are there more complaints registered against credit cards and account opening?

This answer to this could be as more and more people are turning to online payment transactions methods rather than physically visiting the banks hence there were more users who have registered for credit card and account opening services online and there might be heavy network traffic due to which these issues have been arising. There could be another reason where the credit card limit has been exceeded and hence the money transfer got declined by the company or the card being expired, the billing address may be incorrect.

2. Why are more issues registered against managing an account?

With increase in online services no body actually visits banks that often and with easy method of transaction people usually prefer cashless online transaction which results in busy servers as many users will be making a transaction at the same time. This could be one reason for account managing issues.

3. Why are there issues registered against deposits and withdrawal?

This could probably be because of incorrect entry of details like bank account number or atm pins. Internet connections, power failure or too many transactions in one go would also cause an issue.

4. Why are there so many complaints against credit card fraud?

PoS Fraud is one of the biggest fraud these days where small skimming devices are attached to the machines which can hack the card details and can be later on used for making anonymous transactions. This is a major issue which is faced by many users and hence there are so many complaints registered with the firm for the same.

CONCLUSION

Now that we have found the pattern on what kind of issues the consumers are facing and how are they reporting the same we can take relevant actions to minimize these issues and make sure that our customers are served well.

APPENDIX

Google Sheets link for Udemy Courses Dataset

<https://docs.google.com/spreadsheets/d/1nO0xw0CO7C1NiipHcYVGUL62LNxnxtpymY4hIbdsY U/edit?usp=sharing>

Tableau Dashboard link for Udemy Course Overview

https://public.tableau.com/views/UdemyCourseOverview/Dashboard1?:language=en-US&:display_count=n&:origin=viz_share_link

Google Sheets link for Financial Complaints Analysis Dataset

<https://docs.google.com/spreadsheets/d/1QqBL2bG6MoBANFCt5SOr06ku BgL Uesa9VRgS3M1gM/edit?usp=sharing>

Tableau Dashboard link for Financial Complaint Analysis Overview

https://public.tableau.com/views/financialcomplaintsanalysis/Dashboard1?:language=en-US&:display_count=n&:origin=viz_share_link