

Data Analysis Portfolio

I'm very much interested and passionate about the Data Analytics field and I would like to explore it a lot. Whenever I am performing an analysis on the data or cleaning the data, I enjoy and very eager to know the insights from that data. I'm very much passionate about Data Analytics and making business decisions which help company's growth. Also, I'm very much interested in HR Analytics also called as Peoples Analytics and would like to explore it more. Also, apart from technical skills, I also have to keep a Success Mindset, my well-planned daily routine, writing down the things in a timely manner. So, at the end of the day, I would like to see myself as a successful Data Analyst and would like to give my best in this field

Professional Background

Currently, I am pursuing my B. Tech in Computer Science and Engineering at Hyderabad, I'm in my 3rd year. we have been taught a lot of things and I would like to apply them. I would like to explore the Data Analytics field. And I would also like to explore and learn more about HR Analytics also called as Peoples Analytics which I'm very much interested and wanted to explore it specifically and in other areas, domains in Data Analytics too. I took some courses on Coursera regarding Data Analysis and got Certified. And presently working on my Data Analytic skills and working on some case studies, portfolio projects, which I would like to show them on my portfolio. I'm constantly learning every day and building up my skills and would like to give my best in this field. And I would like to update myself by learning SQL, R, Python, Statistical skills, Hypothesis Building, Power BI, Tableau, Data Wrangling, AWS, Google Cloud, Azure cloud, Robotic Process Automation etc. I'm a beginner in these skills and would like to grow high in my future in learning these skills effectively.

My Experience as a Data Analyst:

1. Data Analyst Jul 2021 - Aug 2021

EntryLevel, Virtual

Data Analysis Virtual Experience

- Performed Data Analysis on Udemy Courses Dataset
- Performed Analysis using Excel, Tableau
- Made Business Decisions on Udemy Courses, helped in increasing the revenue growth.
- Storytelling

2. Data Analyst Jun 2021 - Aug 2021

Pantech Solutions, Virtual

- Learnt Storytelling, Pivot table, Automation VBA Macros and Power Query in EXCEL
- Worked on Tableau, PowerBI BI tools.
- Got exposed to Numpy & Pandas, Data Visualization, Kaggle Exploratory
- Learnt Database - MongoDB basics for Data Analysis, Basics of NLP

3. Data Analyst Mar 2021 - Jul 2021

Trainity, Virtual

Mar 2021 - Jul 2021

- Learnt Data Analytics basics such as Excel Pivot Tables, SQL, Python, EDA, Tableau, Hypothesis Testing, AB Testing Statistics etc.
- Worked on personality building, Time Management skills.
- Performed Exploratory Data Analysis on IMDB Data Set
- Performed Root Cause Analysis on Uber Cabs Data Set

4. Quality Analyst Sep 2020 - Jan 2021

Avila IT Group Pvt Limited, Virtual

- Performed Quality check using Excel on Amazon and Walmart Data.
- Performed Product Matching of Walmart and Amazon products.
- Worked on product data of Walmart and Amazon.

Table of Contents

Particulars	Page No.
Professional Background	1
Table of Contents	3
Udemy Project Description	4
The Problem	5
Design	6
Findings	7
Analysis	18
Conclusions	19
HR Analytics Project Description	20
Data Design	21
Findings	22
Data Analysis	31
Conclusions	32
Appendix	33

Udemy Project Description

Situation:

- To present the data on course revenue.
- To understand where opportunities to increase revenue may lie, and track the performance of courses.
- To come into conclusion that is it worth for the Web Development courses to charge more
- So that it will increase the next quarter earnings

Task:

- To check and find whether to charge more or less on Web Development courses.
- To observe and analyse whether Web Development courses are popular courses or not.
- To decide whether in the next three weeks, how they increase their next quarter earnings.
- To present the data on course revenue, and to find out the most popular courses which would help in increasing the company's next quarter earnings.

Action:

- Data Cleaning
- Visualising the Data
- Bringing insights out of that data
- Making business decisions

Result:

- We can say that it is the correct decision to charge more in the Web Development courses.
- Because, they are more popular when compared to the other courses that were provided in the dataset.
- They are very in demand, and has a lot of students feedback along with many were learning that course.
- So, if we increase the price of Web Development courses, it would definitely increase the next quarter earnings.
- It would help in increase their next quarter earnings.

The Problem

The Business problem is to check whether the Web Development courses are in demand or not

To check whether is it worth to charge more on Web Development courses or not

The data is presented and visualised using Pivot Tables and Tableau.

Some of the concepts that we need to concentrate are:

The data regarding different courses from different topics at Udemy

All the information regarding that course

Data regarding the Instructor, his rating, his courses, his experience in teaching, Number of students enrolling for his course etc.

Data regarding the type of the course, Industry, Objective of the course, Content of the course, Course Materials, Additional content and additional support for that course etc

It can be presented by using Excel data presentation techniques, we could use Tableau, using python programming, and also, we could share our insights to the stakeholders by making a PPT.

Complete information regarding the data collection. Where is the data being collected? And who is maintaining that data?

Is the data ethical?

Who are my stakeholders? And what are their requirements specifically? I would note them when they tell their requirements Performance tracking of different courses? On what basis the performance is measured? Is it by rating of the course and instructor? Is it by the course content? Is it by the instructor? Is it by the number of students enrolling into that course? Is it by particular industry or domain that is booming?

How the courses are being charged exactly? On what basis? And how are the prices fixed to those particular courses?

So, these are some of the questions that I would make a note and if I get any other questions into my mind while I am analysing and working with the data, I would include them also.

Design

The data comprises of the information regarding Udemy Courses.

Firstly, I organised that data according to the course domains and

The Steps I took to clean the data:

Removed Duplicates

Removed Blank Cells

Header Naming Conventions

Find and Replace techniques

IF Function

Etc

Visualization tools that I have used:

Pivot Tables in Excel

Tableau Visualizations

Visualization is important step where; we could find many things in it.

I used Excel Pivot Tables in Analysing the Data because, it is easy and can be understood easily.

Also, I used Tableau because, there are many built-in techniques that we could use.

It reduces our work and burden. We just use them according to our need and requirements.

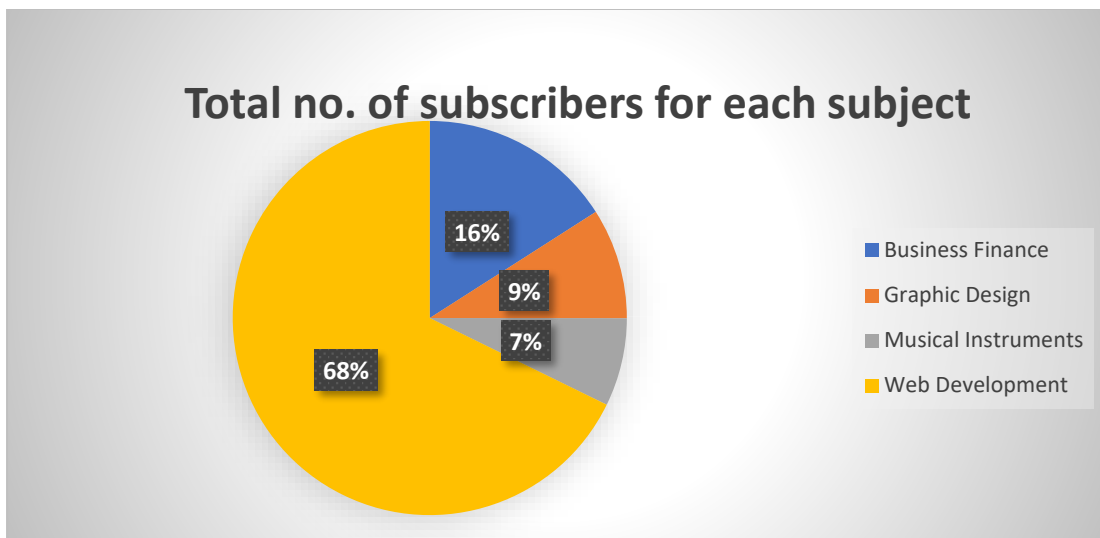
So, they helped me in visualizing.

Findings

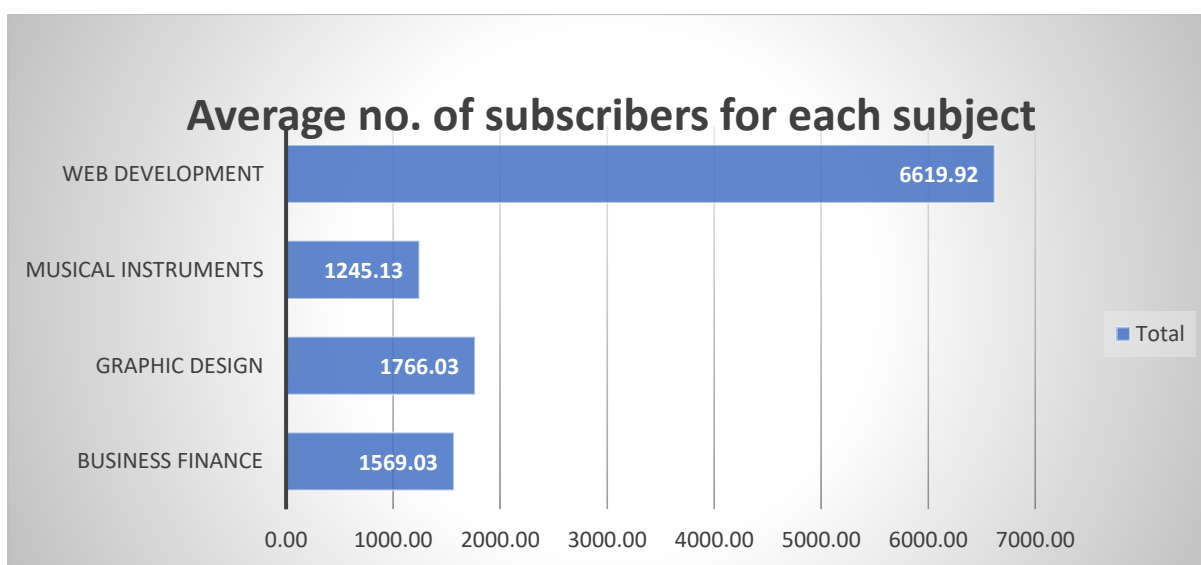
My Visualizations from Excel and Tableau are as follows:

Excel Visualizations and Tables:

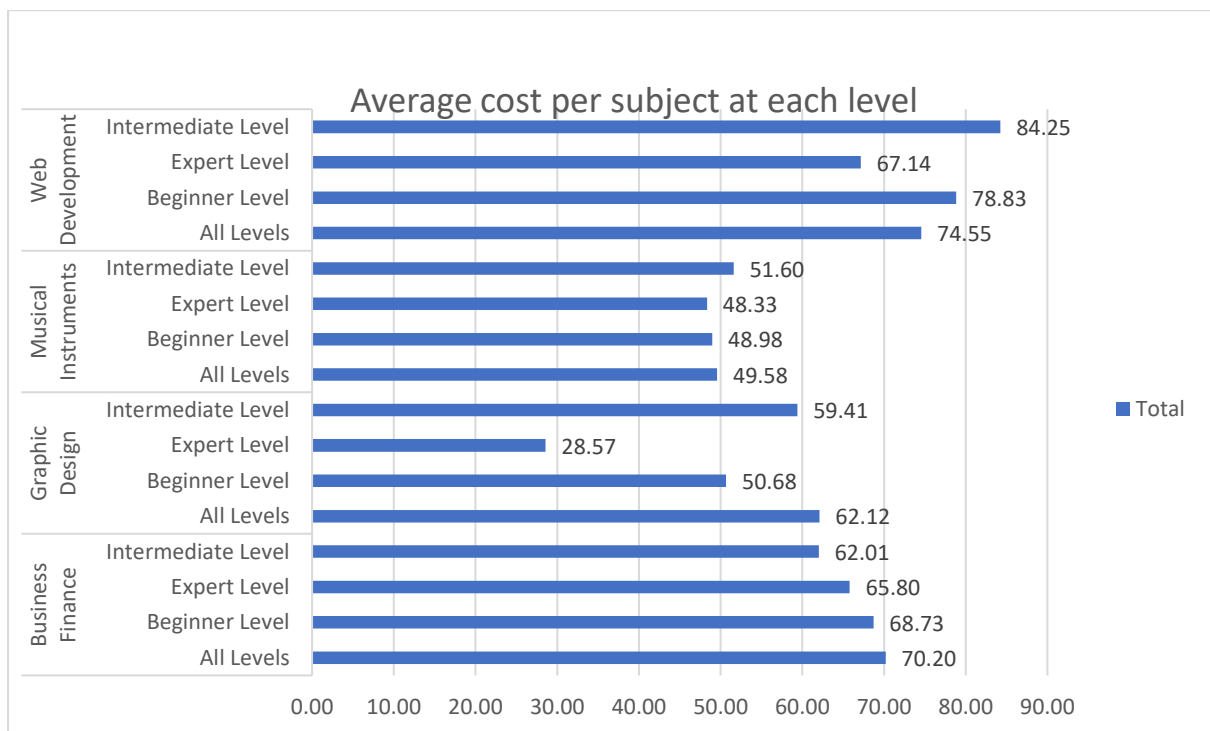
Row Labels	Sum of num_subscribers
Business Finance	1868711
Graphic Design	1063148
Musical Instruments	846689
Web Development	7937287
Grand Total	11715835



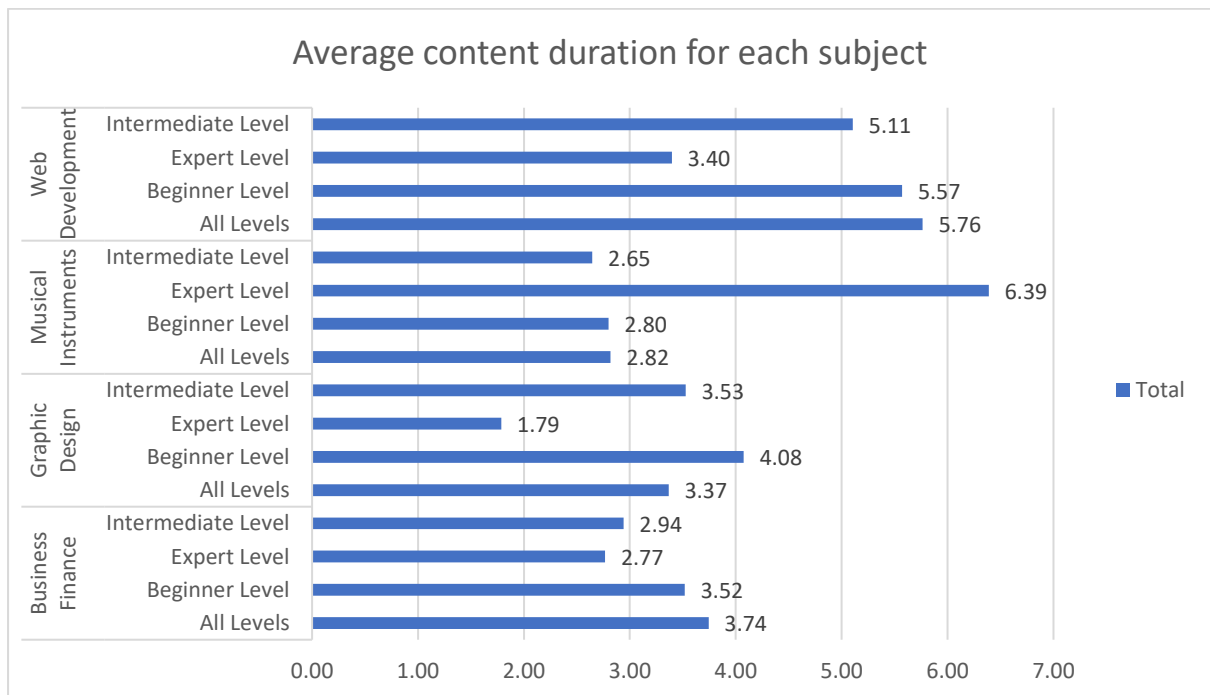
Row Labels	Average of num_subscribers
Business Finance	1569.03
Graphic Design	1766.03
Musical Instruments	1245.13
Web Development	6619.92
Grand Total	3190.59



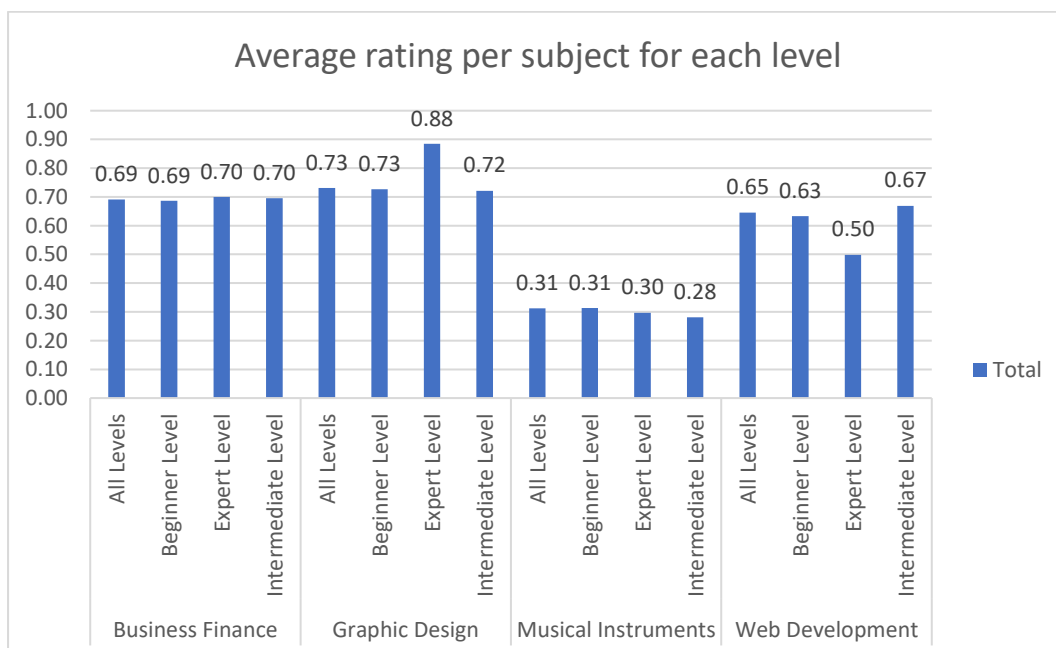
Row Labels	Average of price
Business Finance	68.69
All Levels	70.20
Beginner Level	68.73
Expert Level	65.80
Intermediate Level	62.01
Graphic Design	57.89
All Levels	62.12
Beginner Level	50.68
Expert Level	28.57
Intermediate Level	59.41
Musical Instruments	49.56
All Levels	49.58
Beginner Level	48.98
Expert Level	48.33
Intermediate Level	51.60
Web Development	77.04
All Levels	74.55
Beginner Level	78.83
Expert Level	67.14
Intermediate Level	84.25
Grand Total	66.10



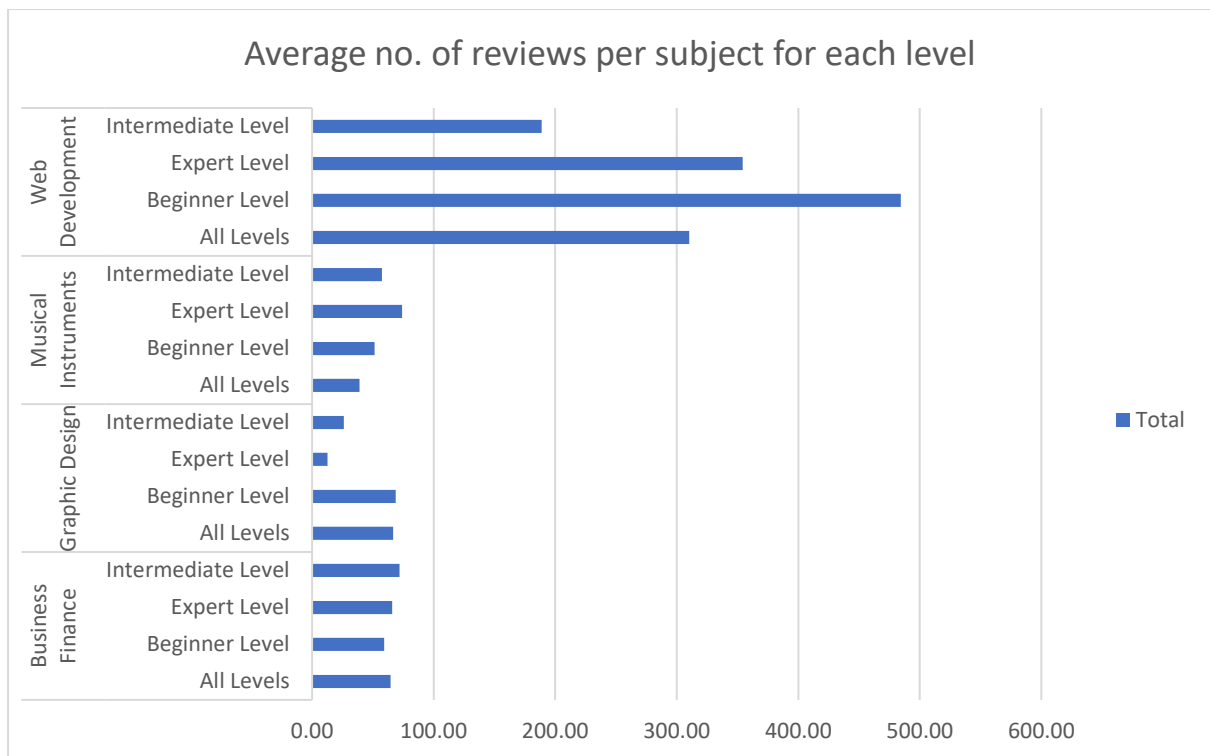
Row Labels	Average of content_duration
Business Finance	3.56
All Levels	3.74
Beginner Level	3.52
Expert Level	2.77
Intermediate Level	2.94
Graphic Design	3.59
All Levels	3.37
Beginner Level	4.08
Expert Level	1.79
Intermediate Level	3.53
Musical Instruments	2.85
All Levels	2.82
Beginner Level	2.80
Expert Level	6.39
Intermediate Level	2.65
Web Development	5.60
All Levels	5.76
Beginner Level	5.57
Expert Level	3.40
Intermediate Level	5.11
Grand Total	4.10



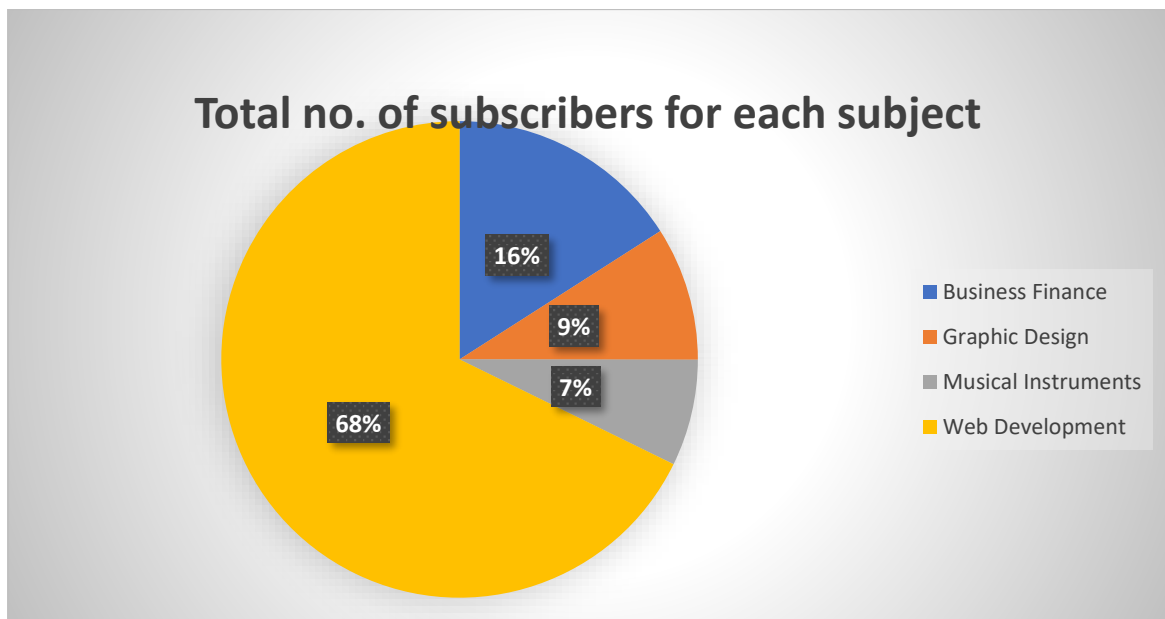
Row Labels	Average of Rating
Business Finance	0.69
All Levels	0.69
Beginner Level	0.69
Expert Level	0.70
Intermediate Level	0.70
Graphic Design	0.73
All Levels	0.73
Beginner Level	0.73
Expert Level	0.88
Intermediate Level	0.72
Musical Instruments	0.31
All Levels	0.31
Beginner Level	0.31
Expert Level	0.30
Intermediate Level	0.28
Web Development	0.64
All Levels	0.65
Beginner Level	0.63
Expert Level	0.50
Intermediate Level	0.67
Grand Total	0.61

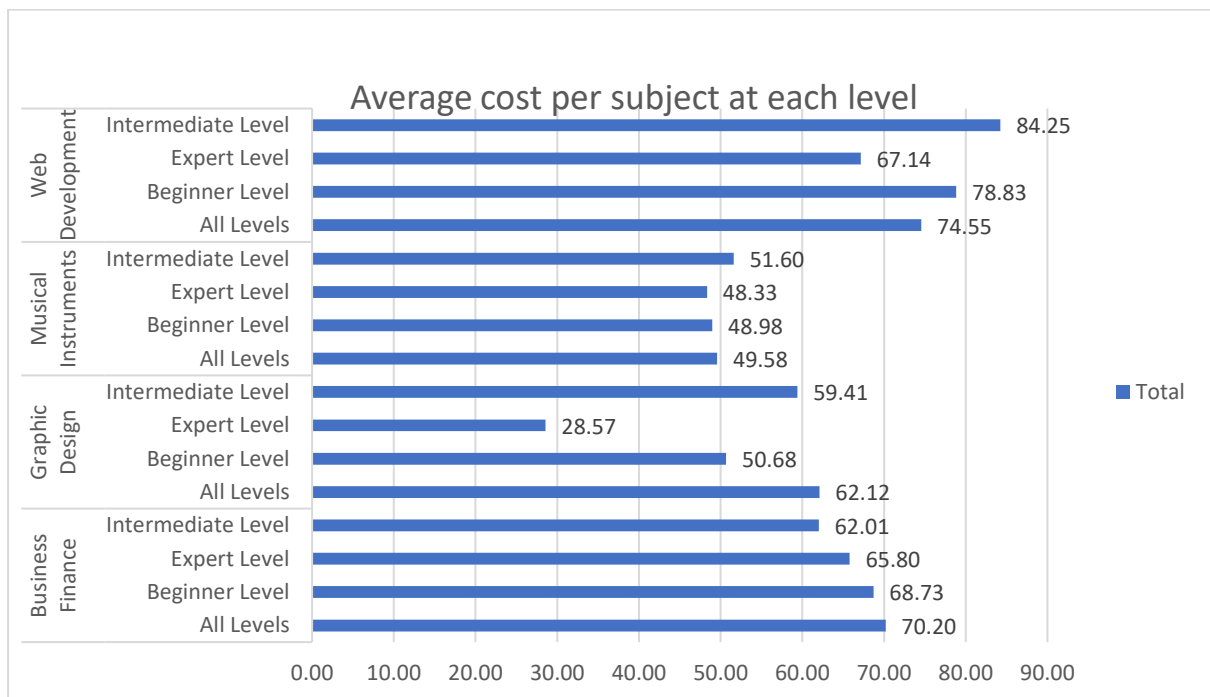
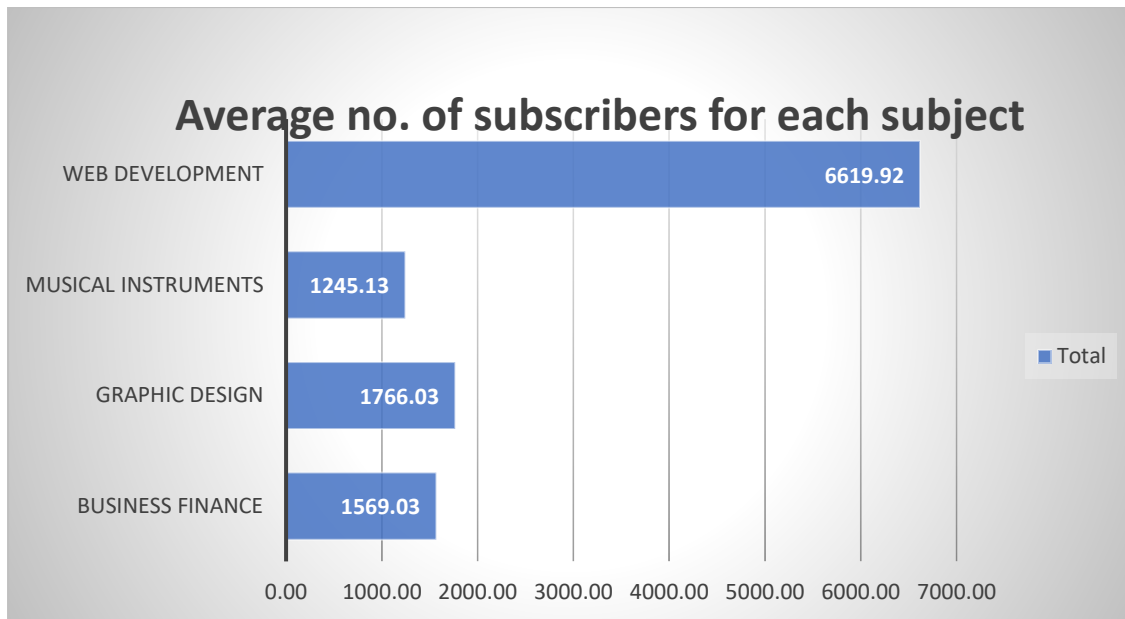


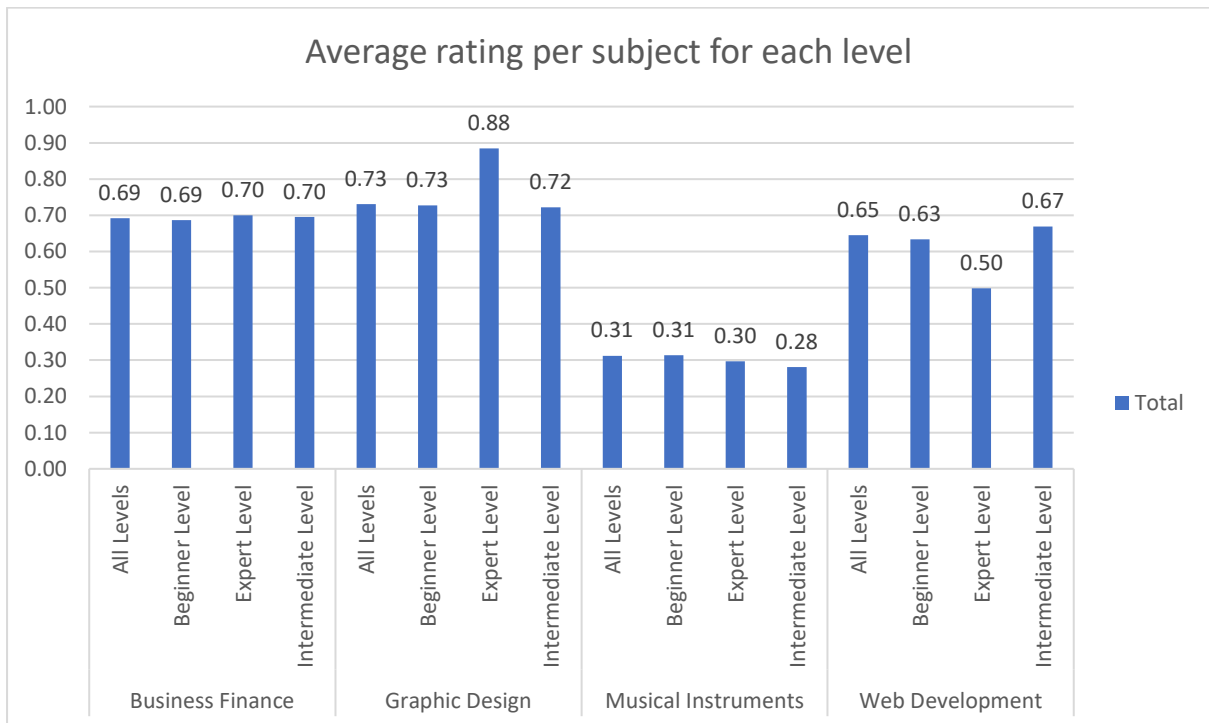
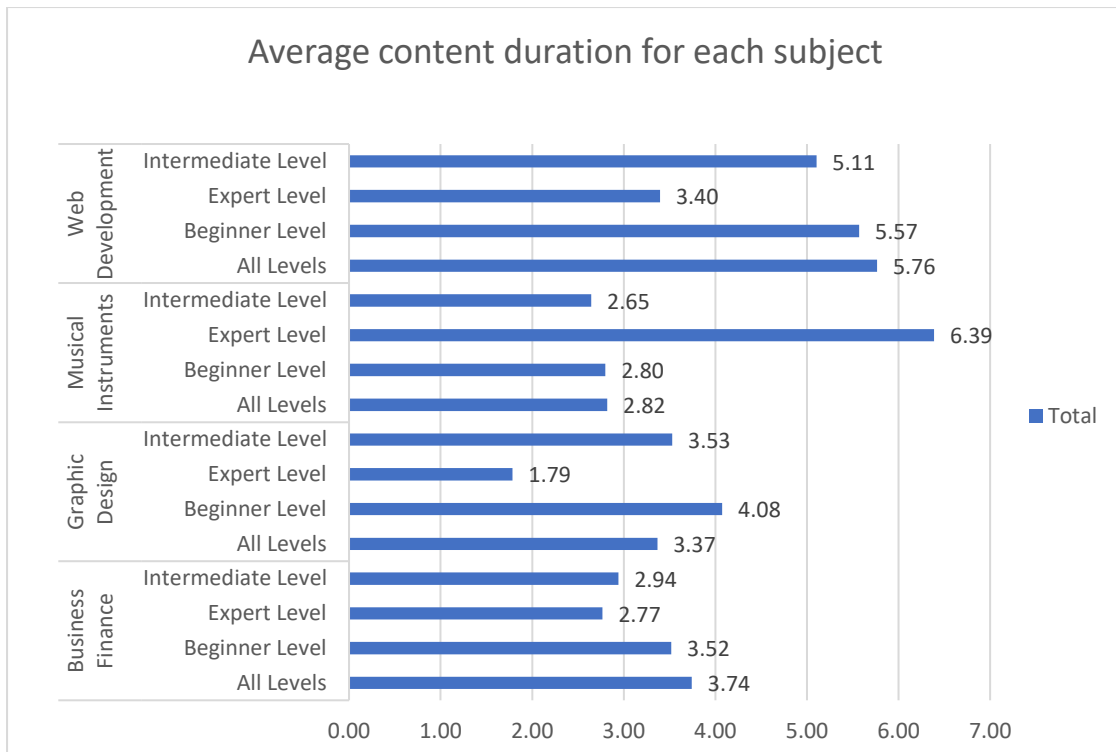
Row Labels	Average of num_reviews
Business Finance	63.73
All Levels	64.66
Beginner Level	59.40
Expert Level	65.84
Intermediate Level	71.85
Graphic Design	61.58
All Levels	66.71
Beginner Level	68.76
Expert Level	12.57
Intermediate Level	26.08
Musical Instruments	46.65
All Levels	39.14
Beginner Level	51.41
Expert Level	74.00
Intermediate Level	57.44
Web Development	358.22
All Levels	310.35
Beginner Level	484.41
Expert Level	354.21
Intermediate Level	188.88
Grand Total	156.37



Excel Findings:







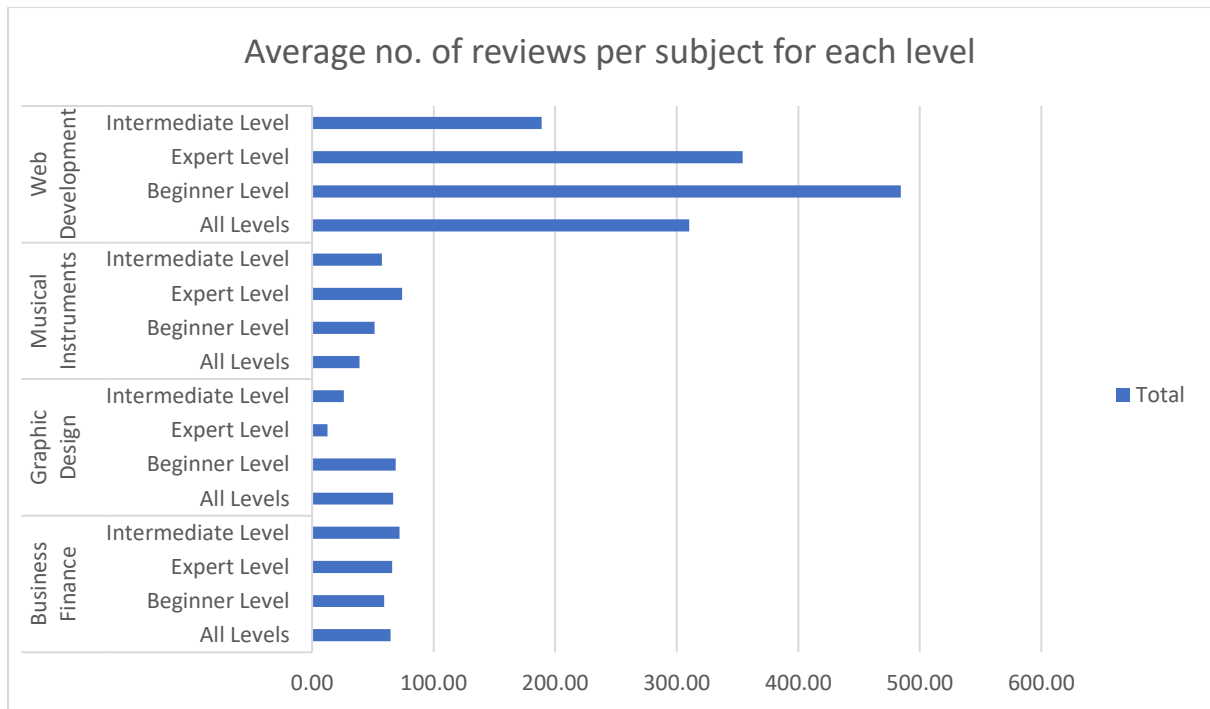
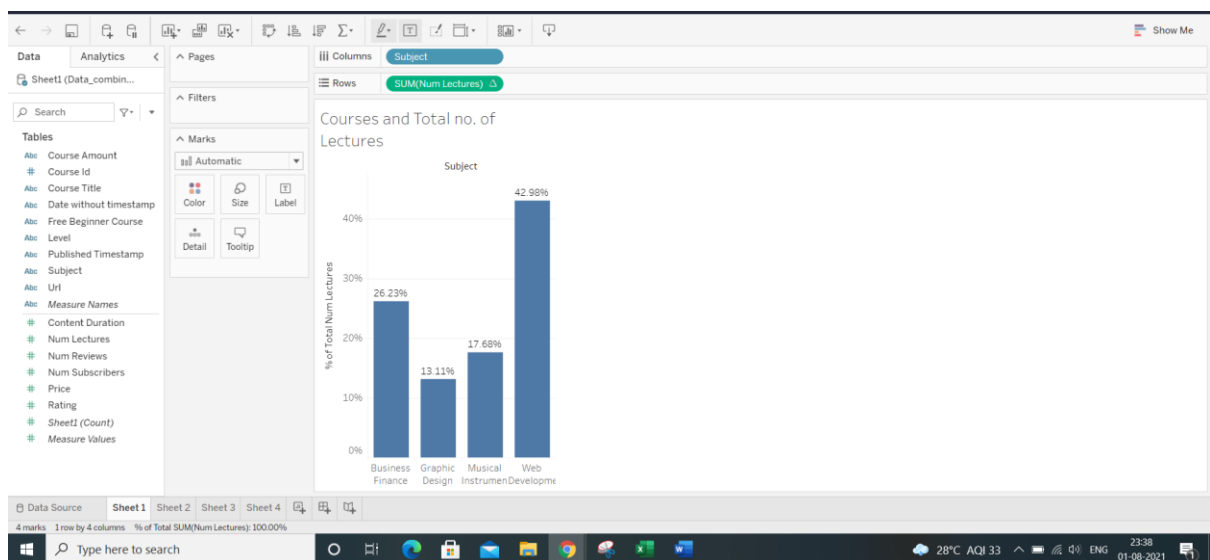
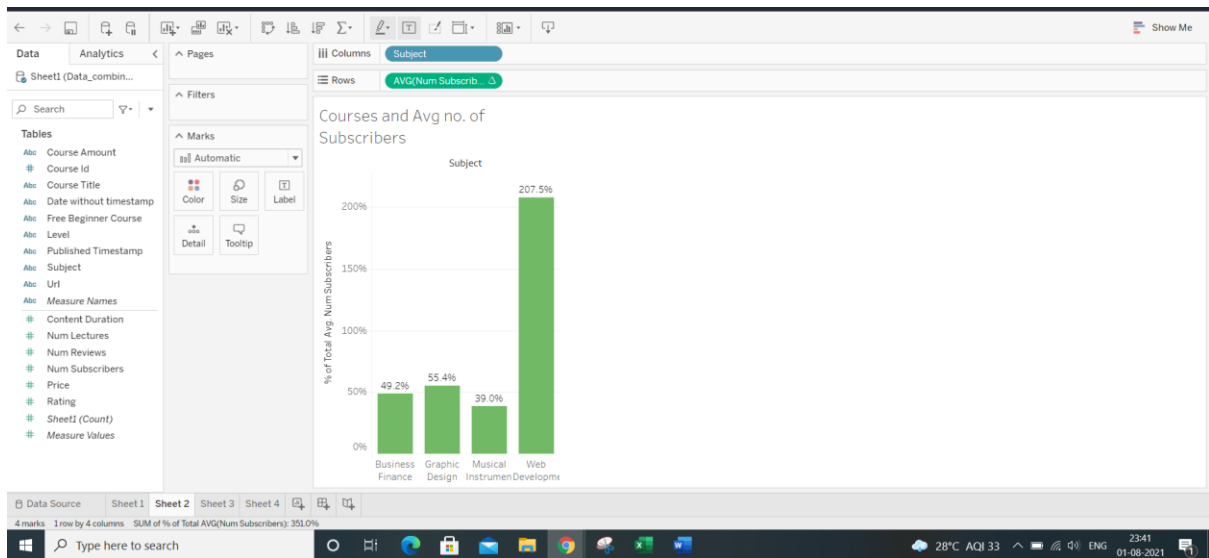


Tableau Visualizations:

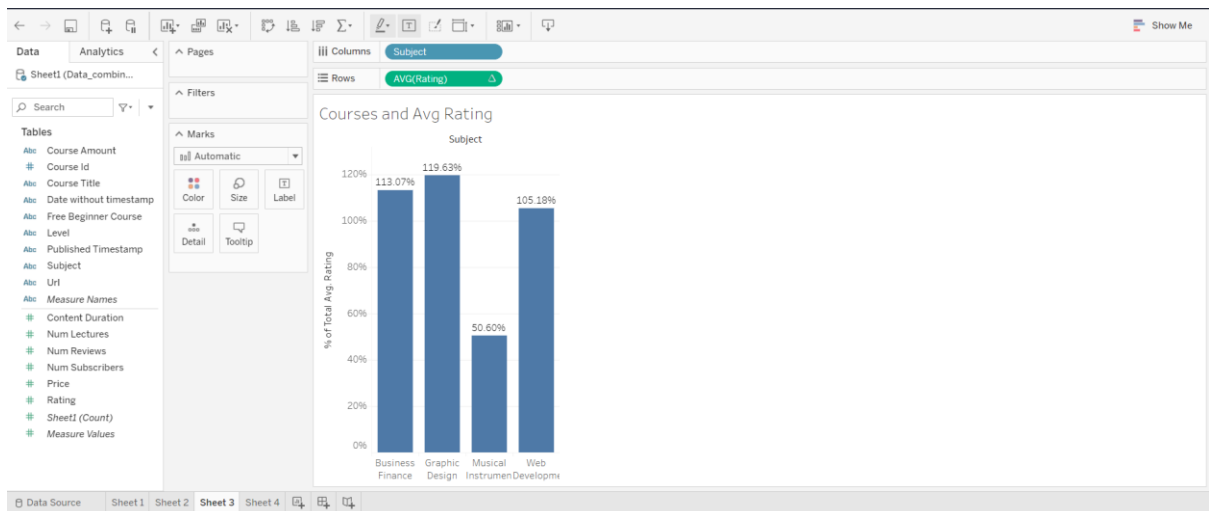
Courses and Total No. of Lectures



Courses and Avg no. of Subscribers



Courses and Avg Rating.



Analysis

5 Why's Analysis:

Why Web Development courses has high no. of Subscribers?

Because the Avg no. of Subscribers is high when compared to other courses.

Why is the Avg no. of Subscribers low for other courses?

Because the Avg cost per subject is one of the reasons.

Why is it like that?

Because the Avg content duration also matters.

Why is that matter?

The Avg rating is different according to different subjects.

Why is the Avg rating different?

The Avg no. of Reviews per subject for each level is the reason.

Conclusions

We can say that it is the correct decision to charge more in the Web Development courses.

Because, they are more popular when compared to the other courses that were provided in the dataset.

They are very in demand, and has a lot of students feedback along with many were learning that course.

So, if we increase the price of Web Development courses, it would definitely increase the next quarter earnings.

It would help in increase their next quarter earnings.

HR Analytics Project Description

The project is about working on the HR data, coming up with some results, HR data has employee data, where we need to analyse that data also analyse the things such as Employee leaving the company, Employee retention etc.

Various factors that would help analysing the HR data, which helps the company to make the decisions and would help to hire the employees who is leaving the company and what are the factors, what are the various things that company need to be concentrate etc

What are the things that company can look to ensure the employee retention?

So, basically, we are analysing the HR data, bringing insights from it.

Data Design

Analysing the HR Data and reasons of Employee retention
Attributes from the Data Set:

satisfaction_level : Employee satisfaction level

last_evaluation : The score that the employee got for his evaluation. That is the review that is given by the company on considering the performance and all other factors of an employee who work in that particular company.

number_projects: Number of projects done by the employee

average_monthly_hours: Average number of hours done with respect to 30 days. Let us consider 30 days here.

time_spend_company: Assuming the number of years spent in the company. The total number of years that the employee spent in the company

Work_accident : Weather or not an accident has occurred in the workplace. An accident may be anything. Like not delivering the output, result sometime etc due to some issues.

left: Weather or not the employee has left the company, it is represented '1' if left and it is represented as '0' if not.

promotion_last_5years: Weather or not the employee has had a promotion it is represented as '1' and it is represented as '0' if not

Department: Department in the company. Like there are various departments.

salary: Degree of salary High, Medium or Low

Looking at this Dataset it has been cleaned thoroughly

Whether the employee will leave or not

Cleaning the data:

There are no null values, blank cells in the data set that is provided

Headers are renamed properly by ensuring the clear and concise names. By adding dashes or underscores in between words to make it easier to parse later on.

There are no duplicates in the data set that is provided

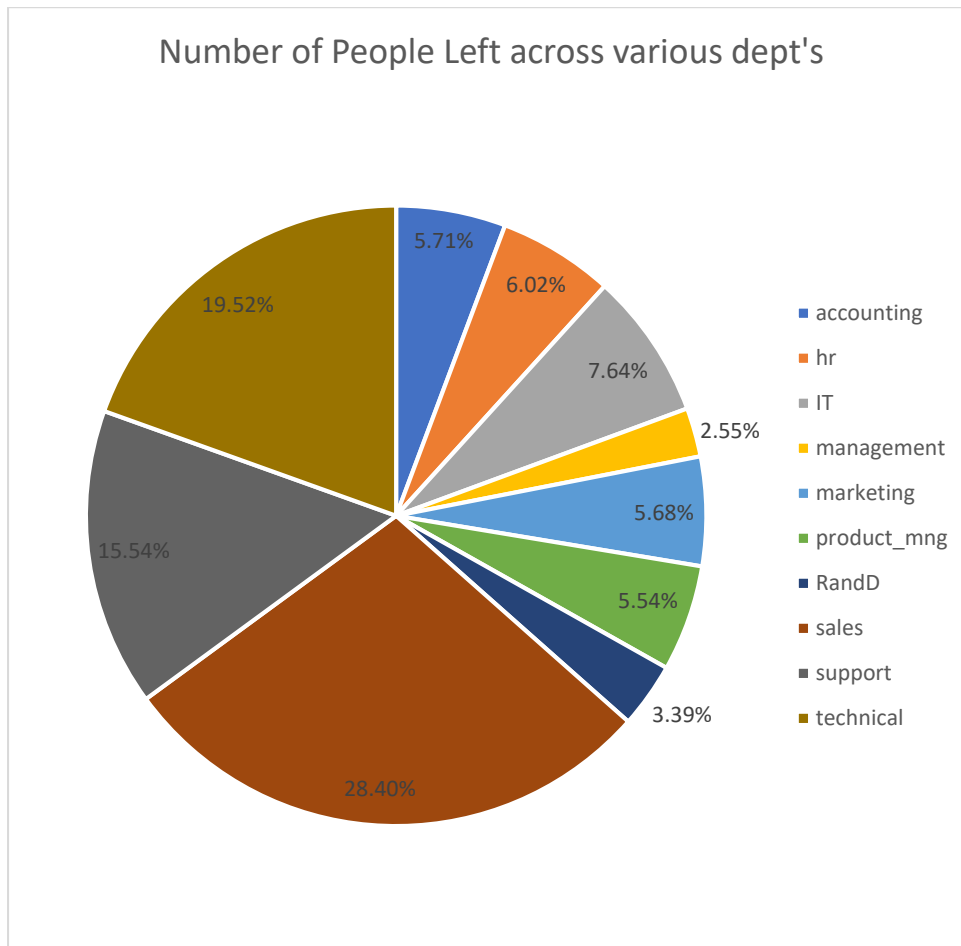
Added a new column for better understanding. Added the Emp_id column for identifying unique employees and Employee Count

Used "Rand()" and "RANK(B2, \$B\$2:\$B\$15000)" to get unique numbers for Emp_id column
Added columns that are required for the analysis

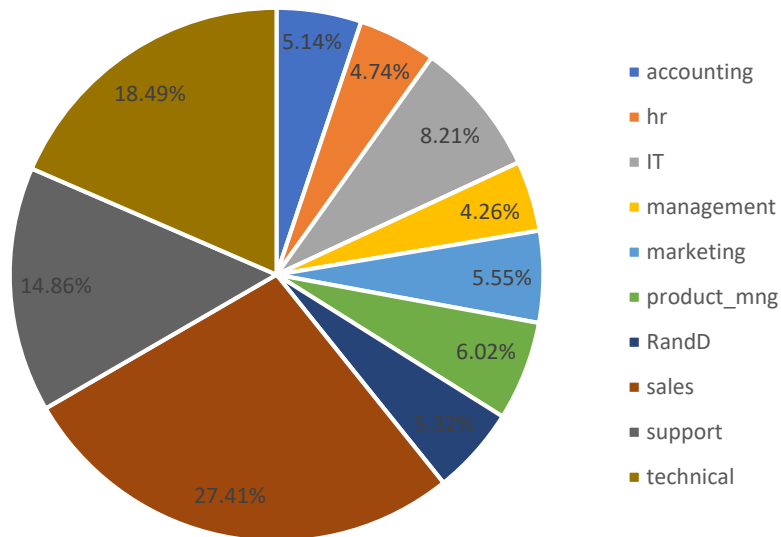
No usage of Find and Replace. Everything looks fine.

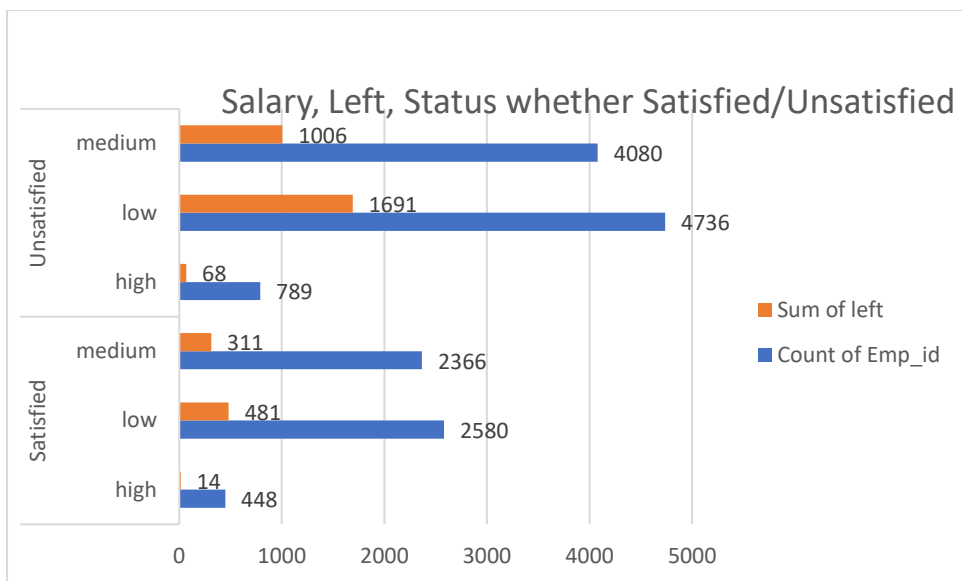
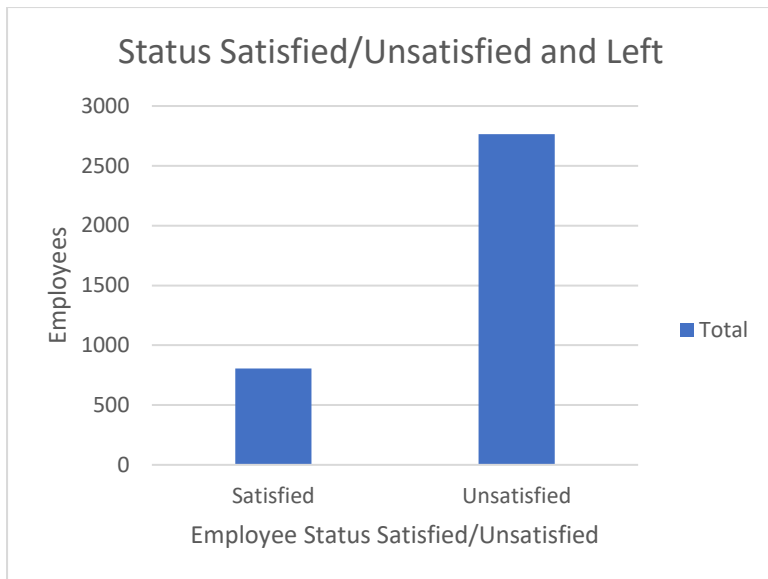
Findings

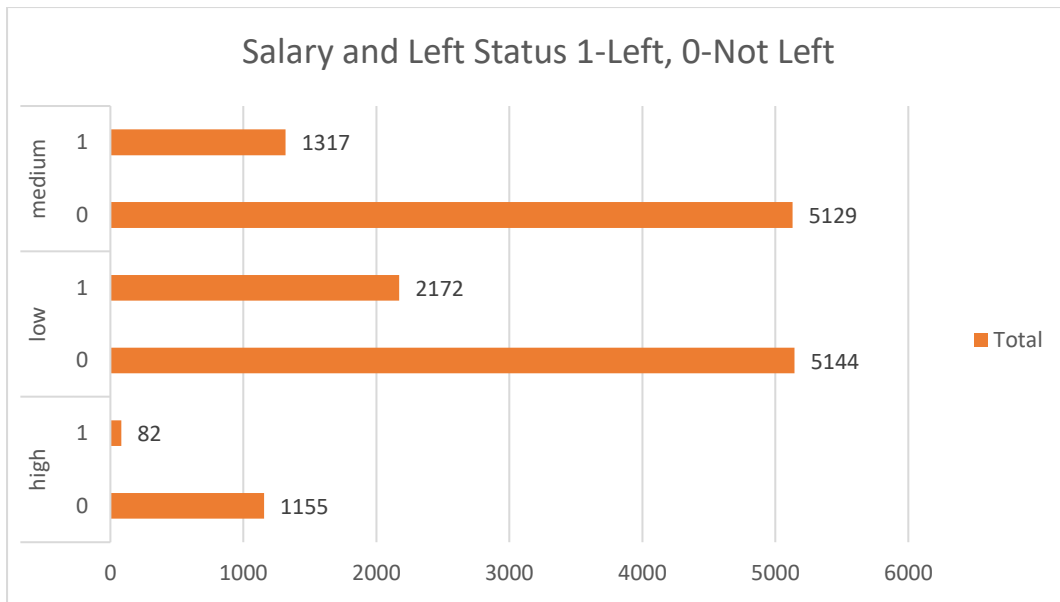
Tableau and Excel Visualizations are as follows:



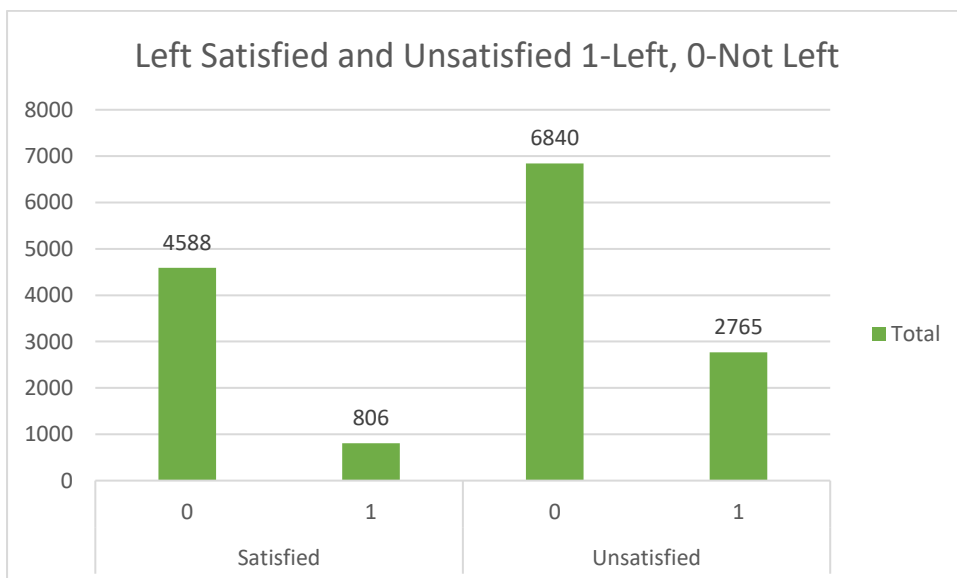
Total Number of Proj's across various dept's



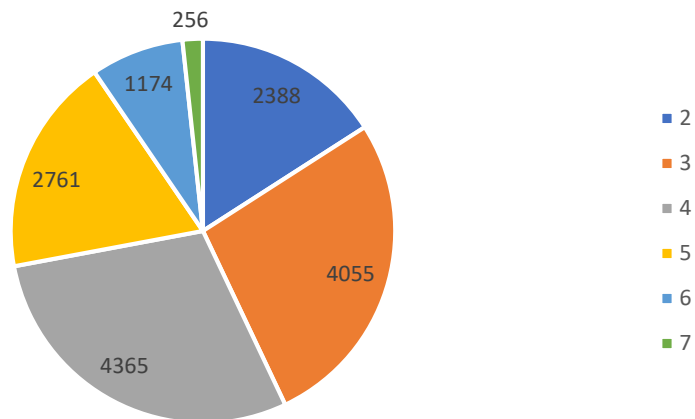




Row Labels	Sum of left
Satisfied	806
Unsatisfied	2765
Grand Total	3571

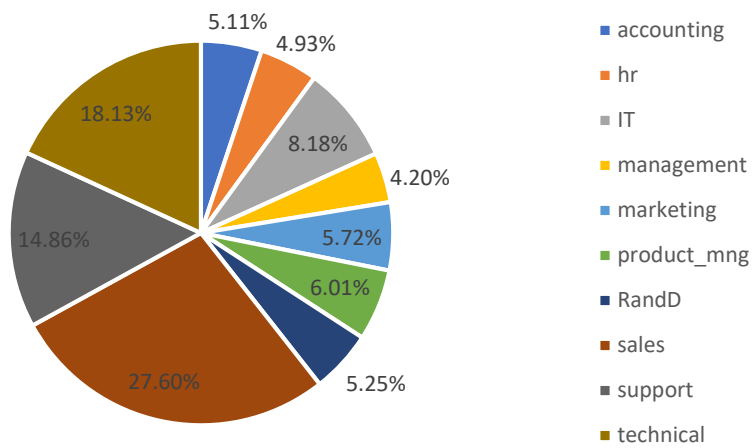


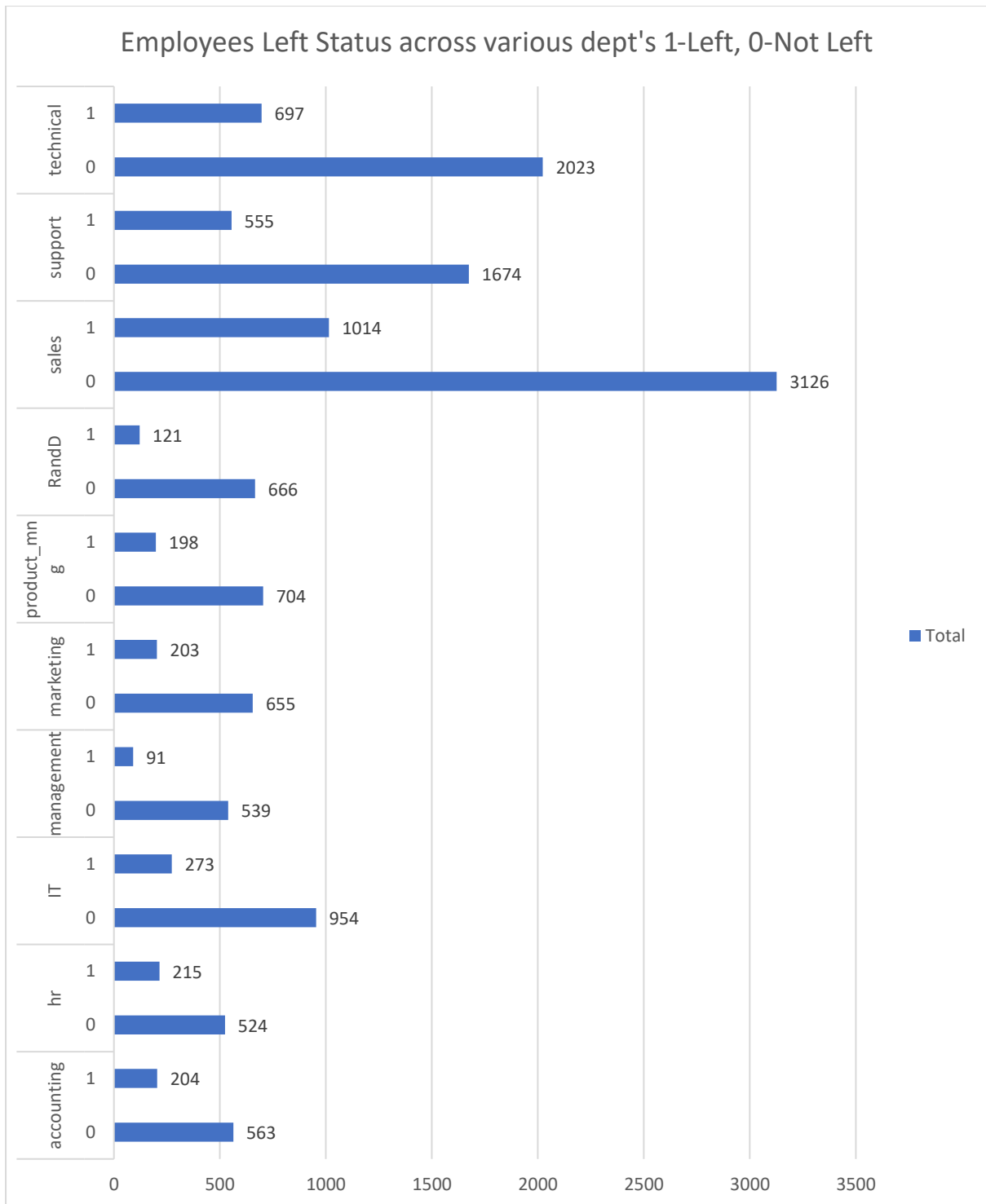
Employees and their Projects

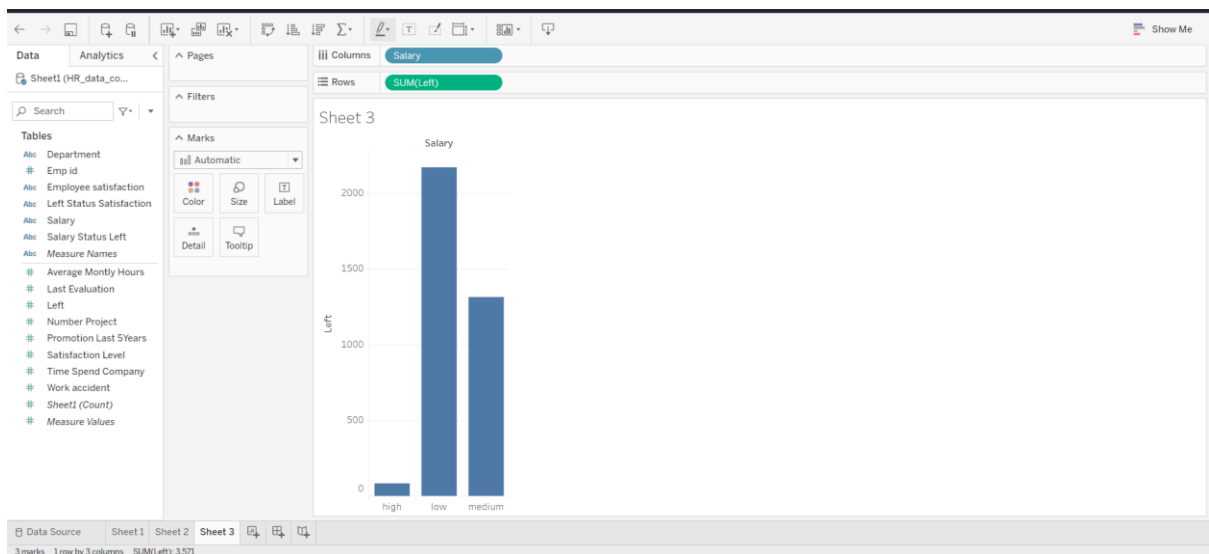
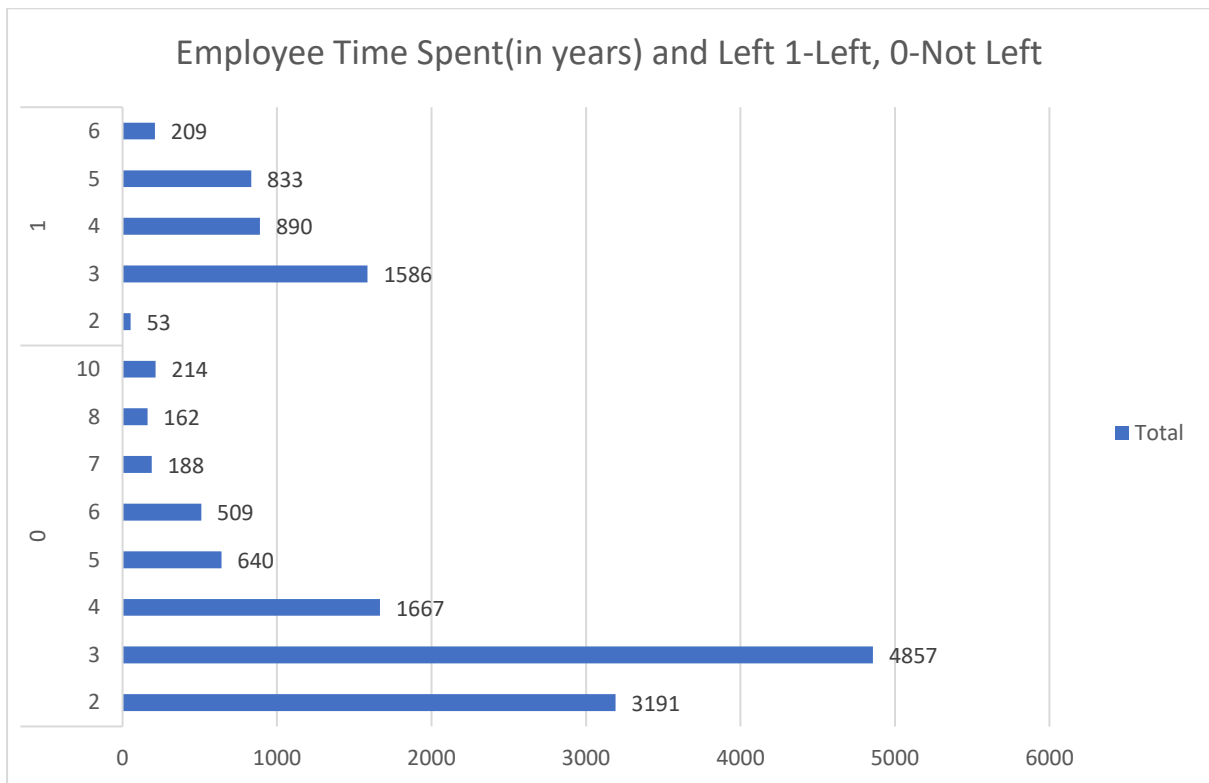


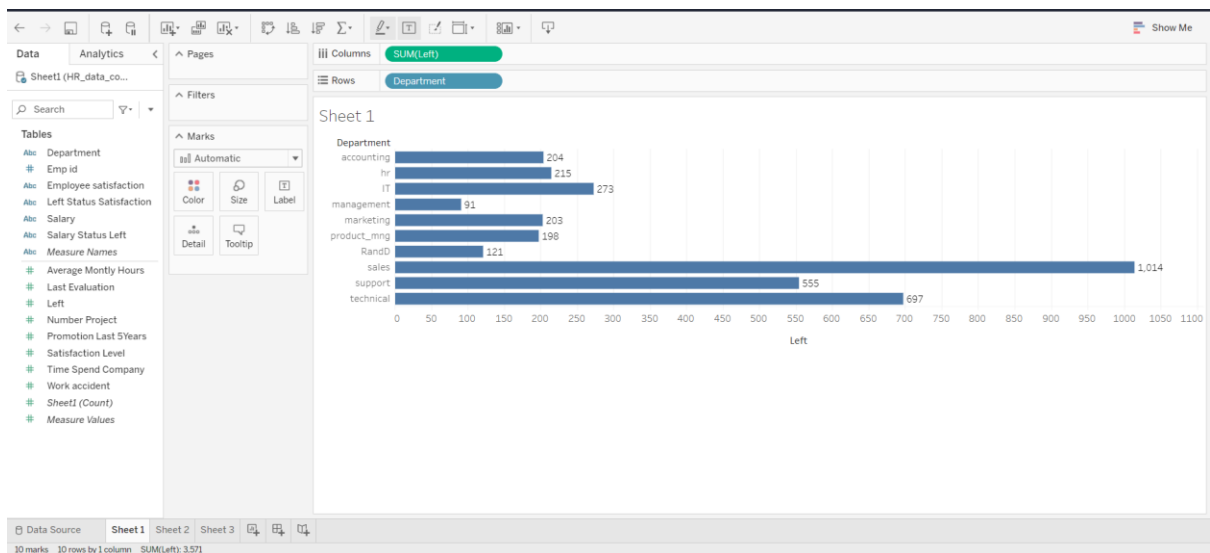
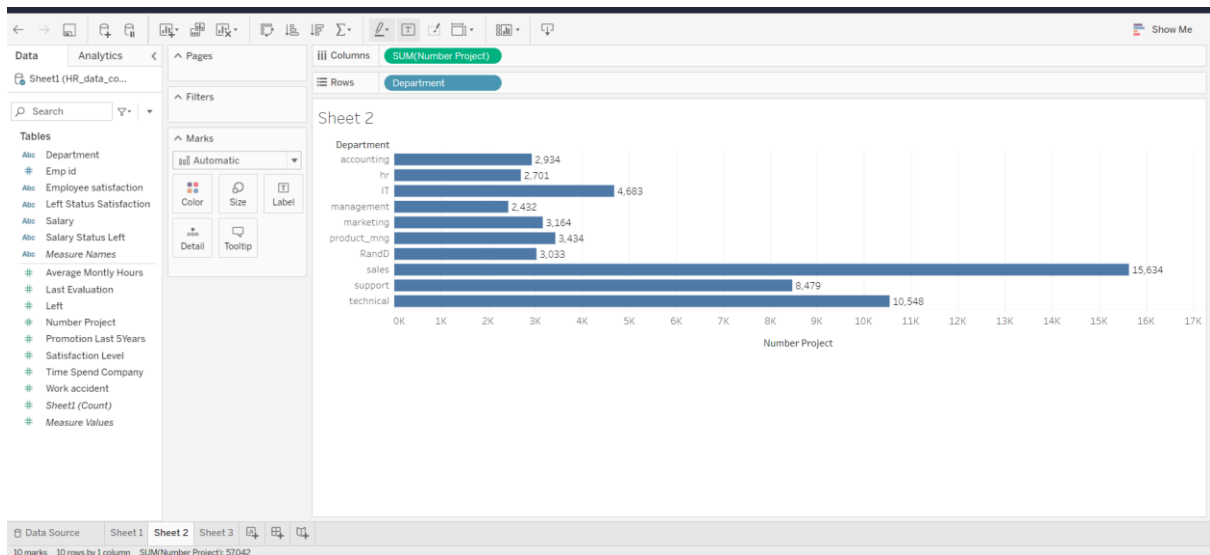
Row Labels	Count of Emp_id
Satisfied	5394
0	4588
1	806
Unsatisfied	9605
0	6840
1	2765
Grand Total	14999

Number of Emp's across various dept's









Data Analysis

Root cause and Why's analysis:

Why are the employees leaving the company?

Why are the several factors important in the decision of employee leaving the company?

Why there is a less satisfaction level?

Why should we consider the factors effecting satisfaction level?

Why do we need to change the internal things?

From the above things we came to the below conclusions.

Conclusions

We Observed that:

- More number of people are working in Technical Department
- Less number of people are working in Management Department
- More number of projects are done by the Technical Department
- Less number of projects are done by the Management Department
- Employees who are unsatisfied left the company more
- Employees who have low salary, and are unsatisfied left the company more
- Many Employees did around 3 to 4 projects on an average
- Most of the Employees left from the Sales Department followed by technical department
- Employees who spent 3 years left the company more
- Employees who have do not have promotion of last 5 years, left the company more.
- The employees who left also seem to have a lower satisfaction level
- Most employees that leave seem to have low or medium salaries. Unsurprisingly most high paid employees don't leave.
- Satisfaction Level seems to be the most important feature.
- Surprisingly the salary seems to be the least important feature to determine whether the employee will leave or not.

Appendix

HR Analytics Project Data Set:

<https://www.kaggle.com/mfaisalqureshi/hr-analytics-and-job-prediction>