Data Analysis Portfolio

Seeking an Entry Data Analyst position with the aim of utilising analytical thinking and relevant expertise to help in the realization of the goals of organisations. Also, to leverage and combine skills and knowledge acquired in Data Analytics, Machine Learning, and Deep Learning to solve problems encountered in different sectors as a whole. I am keen on learning all other germane skills needed to be a more effective Data Analyst, actively working with experts and colleagues to meet this target.

Professional Background/Experiences

MSc PETROLEUM ENGINEERING

Tyumen Industrial University (TIU), Russia, 2020-2022

- Master's thesis on the applications of Data Analysis and Machine Learning in the petroleum industry.

DATA ANALYSIS VIRTUAL LEARNING PROGRAM

Entry Level, July 2021

- Learnt how to determine a problem in any sphere
- Learnt a number of data cleaning methods using Excel/Google Sheets
- Created Pivot Tables to show key insights within dataset
- Learnt basic functionality of Tableau and created data visualisations
- Earned three badges: Critical Thinker, Team Player, and On Time

Table of Contents

Udemy Project Description

Presented with data on different courses from different topics, the situation is that, the quarter earnings of the company has not increased although there is a high patronage of courses like Web Development, and the task is to understand where opportunities to increase revenue may lie, and to track the performance of courses.

The Problem

The problem is to find an avenue to increase quarter earnings and in so doing, it is suggested that Web Development courses need to cost more since they are the most popular courses but, this has to be corroborated with the data given.

I had less than two weeks to present comprehensive findings and conclusions/recommendations from the data given. Data on four course categories were collected and stored in a csv file for analysis.

In order to better understand the problem, further research questions needed to be asked else the suggestion/recommendation would prove unwarranted.

- First, there is the need of a proof from the data collected to verify if indeed Web Development courses are the most popular.
- There is a need to understand the reason behind people patronizing Web Development courses. Its patronage might not be because it is popular but because it is required of people to study at least one web development course before applying for certain job openings and in this case, they might blow off the courses in order to get the certificates awarded at the end of the program. What happens if these courses stop being required in companies that hitherto required them before applying for vacancies?
- Also, the popularity of Web Development courses NOW does not translate into future popularity of these courses so finding a solution geared towards increasing price might prove futile for revenue growth.
- There might be other courses with higher ratings but with low purchase and giving these courses more attention could increase revenue in the new future.

Design

The design stage involved merging all data into one excel sheet and also taking steps to remove duplicates and blank cells from the raw data. It was ensured that headers were clear and concise in order to make parsing easier later on. In addition, there was an error in the input of Web Development in the subject column and this was rectified.

The two most prominent visualisation tools used in this work are Excel and Tableau. Excel is used in a lot of industries and it deals with large amounts of data and succors to find insights. Pivot Tables and Pivot Charts make data visualisation in excel seamless. Tableau on the other hand is one of the industry standards for data visualisations. It is built purposely for this task and as such is embedded with a lot of functions to make drawing insights from data very simple and interesting. It works well with very large data sets as well.

Findings

Findings are linked to the Excel Pivot Charts. I had seven (7) major findings and they are as follow:

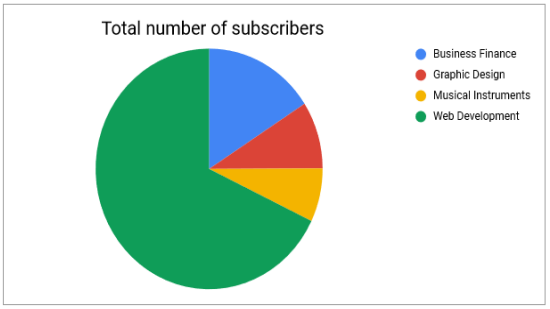
Finding 1

Web Development courses make up approximately 2/3 of the total subscriptions of all courses followed by Business Finance, Graphic Design, and Musical Instruments courses make up the least subscribed courses with a percentage less than 8.

Figure 1: Total Subscriptions by Category

Figure 1

Subject	Total Subscribers
Business Finance	1868711
Graphic Design	1063148
Musical Instruments	846689
Web Development	7981935
Grand Total	11760483



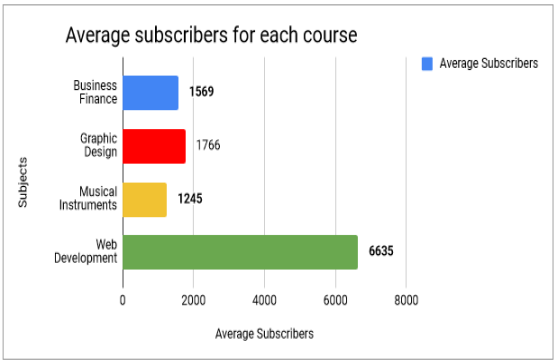
Finding 2

Although Web Development courses have the highest average price at each level, they are still the most subscribed courses indicating that the price of courses plays a trivial role when learners decide to purchase Udemy courses.

Figure 2: Average subscriptions for each course

Figure 2

Subject	Average Subscribers
Business Finance	1569
Graphic Design	1766
Musical Instruments	1245
Web Development	6635
Grand Total	3199.26



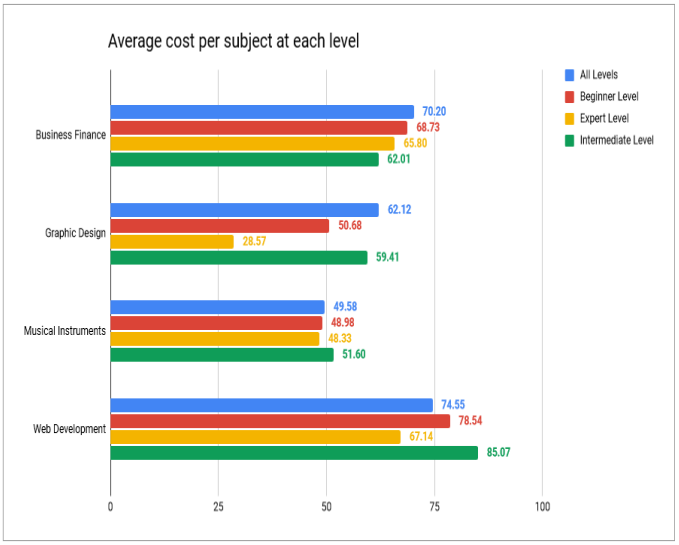
Finding 3

This supports our findings in point two that subscribers are not held back by the prices of the courses, they readily purchase whatever they are interested in.

Figure 3: Average cost per subject at each level

Figure 3

AVERAGE of price	Level				
Subject	All Levels	Beginner Level	Expert Level	Intermediate Level	Grand Total
Business Finance	70.20	68.73	65.80	62.01	68.69
Graphic Design	62.12	50.68	28.57	59.41	57.89
Musical Instruments	49.58	48.98	48.33	51.60	49.56
Web Development	74.55	78.54	67.14	85.07	77.04
Grand Total	66.75	65.24	58.02	66.94	66.12



Finding 4

Web Development courses have the highest average content duration asides having the highest subscribers and highest average price at each level. This is followed by Business Finance and Graphic Design courses with average content durations of 3.56 and 3.59 respectively.

Figure 4: Average content duration of each course

Figure 4

Subject	Average Content Duration
Business Finance	3.56
Graphic Design	3.59
Musical Instruments	2.85
Web Development	5.59
Grand Total	4.10

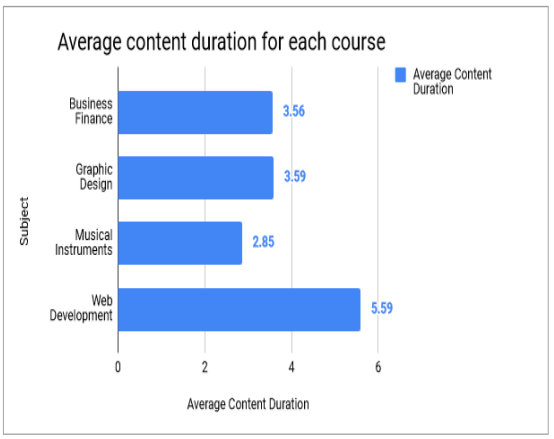
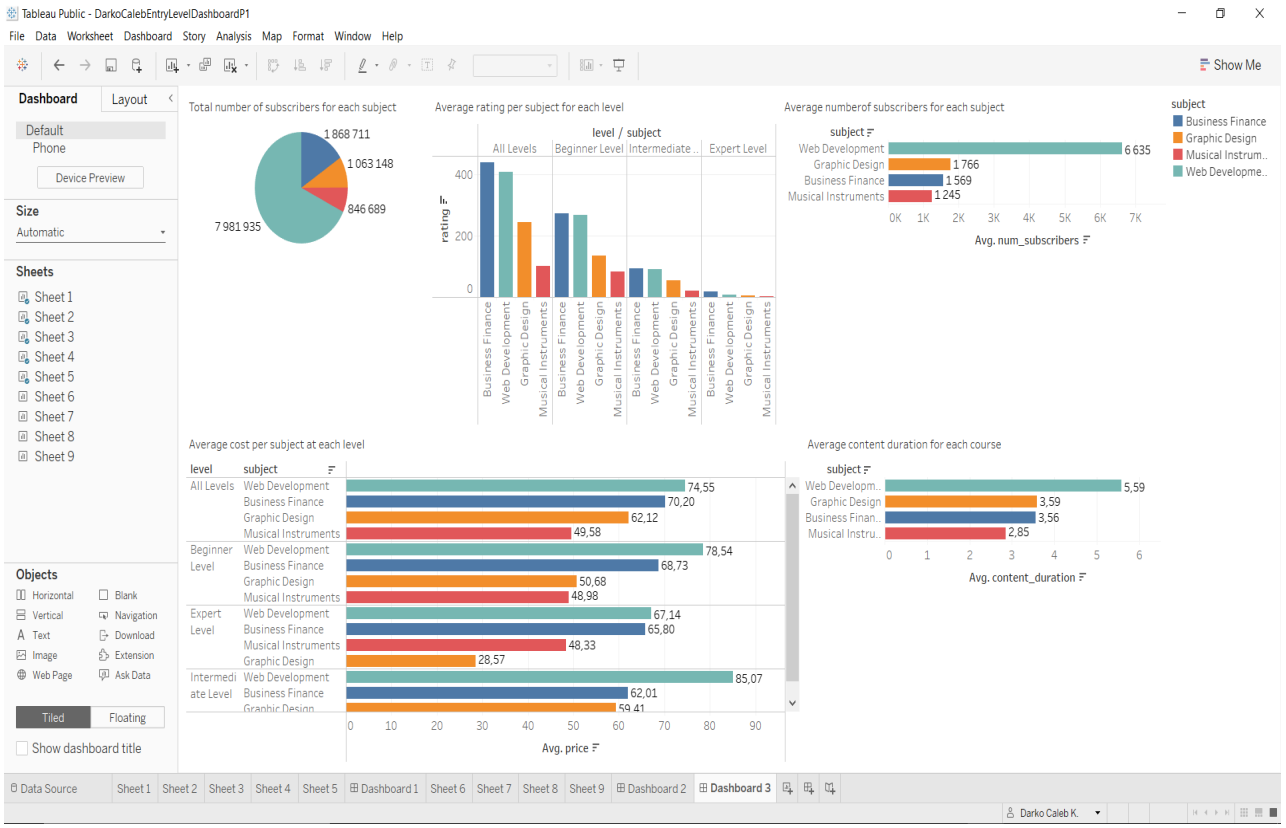
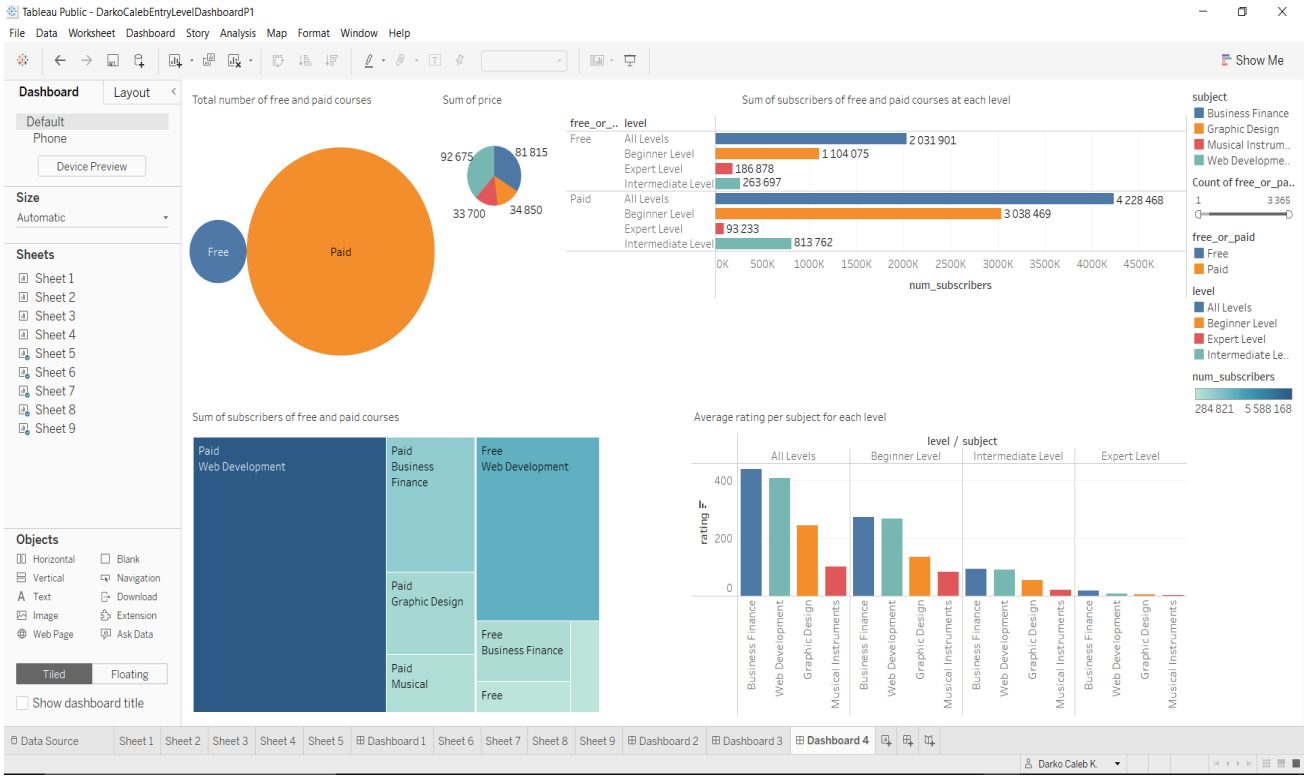


Figure 10: Tableau Visualisations





Analysis

Problem:

The business problem is to draw insight from the data and come out with the best way to increase the next quarter earnings of the company.

Why?

There is an increase in patronage of online courses.

Why?

People find learning online more accessible.

Why?

The advent of covid-19 has made in-person learning difficult and not salubrious.

Why?

There is restriction on the movement of people in countries in order to prevent the spread of covid-19.

Why?

This is to make sure that lives are preserved.

Recommendations

1. Increase the prices of Business Finance and Graphic Design courses after revamping them to learners' preference to help increase revenue.

2. Boost revenue from Business Finance and Graphic Design courses by increasing their average content durations.

3. Business Finance and Graphic Design courses have the highest average rating per subject at each level and this is more the reason why we will need to pay more attention to these courses in order to boost our revenue for the next quarter

4. Monetize a percentage of the Free Web Development courses to increase revenue.

5. It is evident that although Business Finance courses have 15.9% of the total subscribers, they still draw in 33.7% of the total revenue, which is 4.4% short of the revenue generated by Web Development courses (with the highest subscribers of 67.9%). This is yet another reason why Business Finance should be given more attention if we want to increase revenue for the next quarter.

Titanic Project Description

The task here is not to use machine learning to create a model that predicts which passengers survived the Titanic shipwreck but to use data analysis to find out the Gender with the highest survival rate and the factors that brought about the increase in the survival rate of this group.

Presented with data on different courses from different topics, the situation is that, the quarter earnings of the company has not increased although there is a high patronage of courses like Web Development, and the task is to understand where opportunities to increase revenue may lie, and to track the performance of courses.

The Problem

This is the titanic training data set from Kaggle. The problem is to find out the Gender with the highest survival rate and the factors that brought about the increase in survival rate of this group.

I had less than 3 days to present comprehensive findings and draw conclusions from the data given. As mentioned above, the task here **is not to use machine learning** to create a model that predicts which passengers survived the Titanic shipwreck but to **use data analysis** to find out the Gender with the highest survival rate and the factors that brought about the increase in the survival rate of this group

I needed to address questions like:

- what gender was more likely to survive
- what factors helped them survive

Design

There was not a lot of work done with respect to data wrangling. That notwithstanding, the headers were rewritten for easy parsing and reading. The cabin column was deleted because it had a lot of missing data. In addition, the mean age was applied for the few cells with missing ages. Two cells with missing "embarked" data were deleted.

More Information about the data:

ticket_class: 1 - first class, 2 – second class, 3 – third class

embarked: C = Cherbourg, Q = Queenstown, S = Southampton

Excel and Tableau were used for data analysis and visualisations.

Findings

Four (4) of my major findings are presented below:

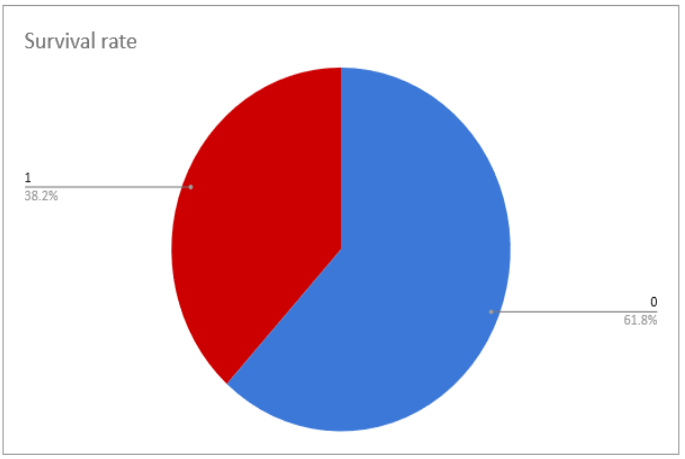
Finding 1

This tells us the number of people who survived out of 889 passengers. 38.2% survived and this comprises people of the male and female genders.

Figure 1: Survival rate

Figure 1

survived	Survival rate
0	549
1	340
Grand Total	**889**



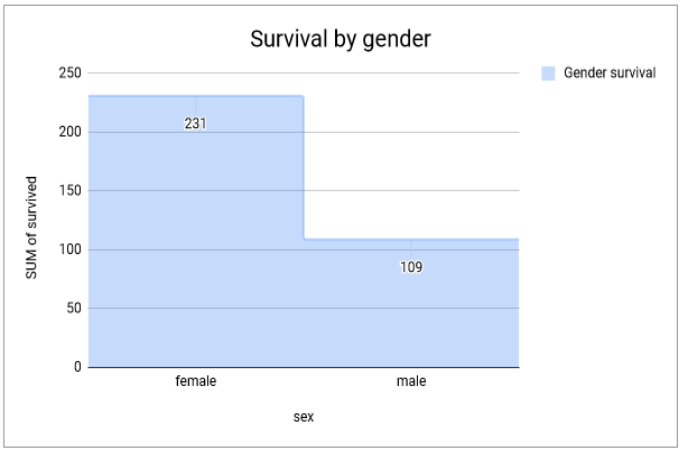
Finding 2

This throws more light on the gender with the highest number of survivors. It is evident that out of the 38.2% survivors of the crash, more than half were females.

Figure 2: Survival by gender

Figure 2

sex	Gender survival
female	231
male	109
Grand Total	**340**



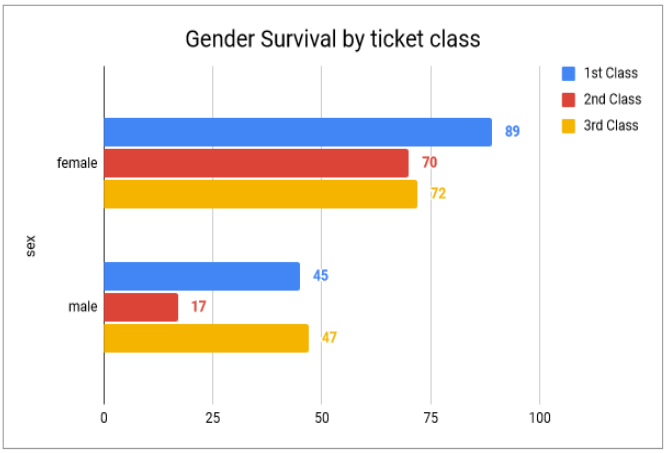
Finding 3

This explored the number of survivors based on the class of the ticket purchased. We find out that most of the survivors were people with first class tickets with women in this category making up twice the number of men.

Figure 3: Gender survival by ticket class

Figure 3

Gender survival	ticket_class			
sex	1	2	3	Grand Total
female	89	70	72	231
male	45	17	47	109
Grand Total	**134**	**87**	**119**	**340**



Finding 4

This seeks to find out the average age at which one is more likely to survive the crash. It is noted that females who survived had an average age of 28 years and males had an average age of 31 years. In general, the average age for survival on the titanic is 30 years.

Figure 4: Average age with survival rate

Figure 4

sex	AVERAGE of age	SUM of survived
female	28	231
male	31	109
Grand Total	**30**	**340**

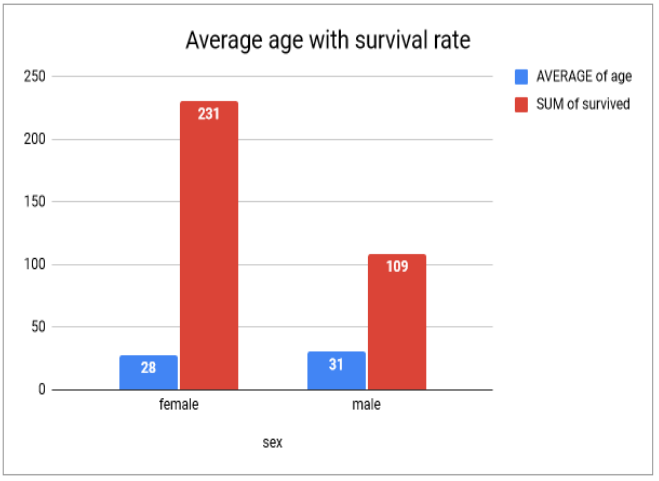
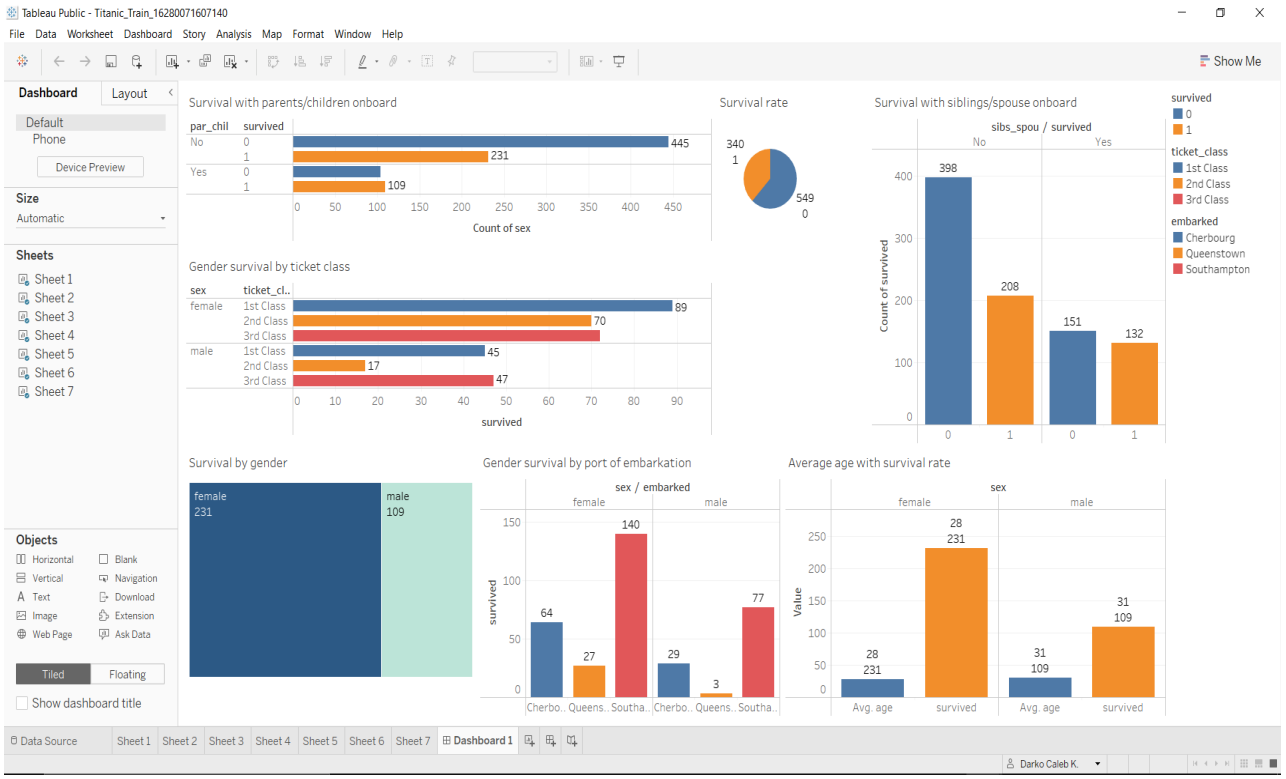


Figure 5: Tableau Visualisation



Analysis

Problem:

The problem is to find out the Gender with the highest survival rate.

Why?

To help know the factors that led to an increase in the survival of this group.

Why?

To understand and give weight to these factors when planning a voyage or anything akin to the experience of the titanic.

Why?

To ensure safety of people.

Why?

To protect human and capital resource.

Why?

To ensure development, growth, and progress in communities, societies, and nations as a whole.

Conclusions

1. Being a female increased one's chance of survival on the titanic.
2. People with first class tickets had a better survival rate especially if a female.
3. Be between the ages of 28 and 31 inclusive to increase your chance of survival.
4. Females who boarded the titanic at Southampton had the highest number of survivals than those who joined in other ports
5. Ultimately, to have had a very high probability of survival on the titanic, one had to be a female in her late 20s, with a first class ticket, and should have boarded the ship at the Southampton port.

Appendix

Google Sheets Data Set for Udemy Project

https://docs.google.com/spreadsheets/d/12GpnWfGyHGA6DlHpoILgrJW2x0j-UP78gxpuTc0QKbU/edit?usp=sharing

Tableau Data Visualisation for Udemy

https://public.tableau.com/views/DarkoCalebEntryLevelDashboardP1/Dashboard3?:language=en-US&:display_count=n&:origin=viz_share_link

https://public.tableau.com/views/DarkoCalebEntryLevelDashboardP2/Dashboard4?:language=en-US&:display_count=n&:origin=viz_share_link

Google Sheets Data Set for Titanic Project

https://docs.google.com/spreadsheets/d/1tgntcfXD7EzffDrU183aqPvoT53BrbXli1Wrh4KzhMY/edit?usp=sharing

Tableau Data Visualization For Titanic Project

https://public.tableau.com/views/Titanic_Train_16280071607140/Dashboard1?:language=en-US&:display_count=n&:origin=viz_share_link