Objectives

- Create a simple dashboard in IBM Watson Studio

Dashboards are an important tool that enable insight into a dataset. The majority of visualization code in a data science project should exist as reproducible scripts, but it is sometimes useful to import tools into the data science workflow. In the case of dashboarding many of the tools available use a graphical interface that offer many ways to tell the story of your data or results. The tool in this tutorial uses IBM Cognos Dashboards, which is available as an embedded service in IBM Watson Studio. This tutorial shows how a dashboard can be used as an investigative tool into a data set.

- [IBM tutorial for interactive dashboards on Watson Studio](#)

The dataset

The original dataset is the classical titanic dataset. We have taken the liberty to modify some of the features to mimic a dataset that more aligns with the AAVAIL company. Let's us walk through the whole process of creating a dashboard as an investigative. AAVAIL gave away free trials in three separate markets (corresponding to the ports where passengers embarked). For each passanger there are a number of features. Here survival indicates whether or not a customer made it to point of becoming a paid subscriber.

1. Customer Id: Unique id given to each customer
2. Is_Subscribed: If a customer is subscribed or not.
3. Tier of Free Trial: There are 3 tiers of free trial, 1st, 2nd and 3rd.
4. Name
5. Sex
6. Age
7. The Subscribtion number
8. The Market: This describes three possible markets that the free trial was carried out in. Three possible values S,C,Q

Types of features

- Categorical - Tier_Free_Trial, Sex, Market, Is_Subscribed
- Continuous - Age, Customer Id, Subscription_Number
- Alphanumeric - Name

Let's load the dataset in IBM Watson Studio. Download the aavail-customer-data.csv file below to get started:

aavail-customer-data.csv

Data summary

Once summarized you can make the following observations:

1. There are a total of 891 customers in our training dataset.

2. Since the Is_Subscribed column has discrete data, the mean gives us the number of people survived from 891 i.e. 38%.

3. Most people belonged to Tier_Free_Trial = 3

Taking a look at our categorical features we find that:

1. Most of the subscribed customers were men.

2. Market has three possible values with most customers having maximum number of streams.

3. Names of all customers are unique.

Building dashboards

The AAVAIL dataset provides observations for each customer and subscribers outcome. The problem statement entails predicting whether a customer would be subscriber or not given the features such as customers, market, gender, age, number of streams, Is_Subscribed and others. In this dashboard we will be analyzing each of the stated features with respect to Is_Subscribed outcome based on the training dataset. Analyses based on data visualization can be an effective tool to provide insight into conditional probability relationships between features and the outcome variable.

Now that we have our observations, let's create a simple dashboard on the IBM Watson Studio. To create the IBM Watson Dashboards follow the instructions in here: https://developer.ibm.com/tutorials/create-interactive-dashboards-on-watson-studio/.

The first visualization we will be creating is to calculate the subscriber percentage and subscriber vs free trial.
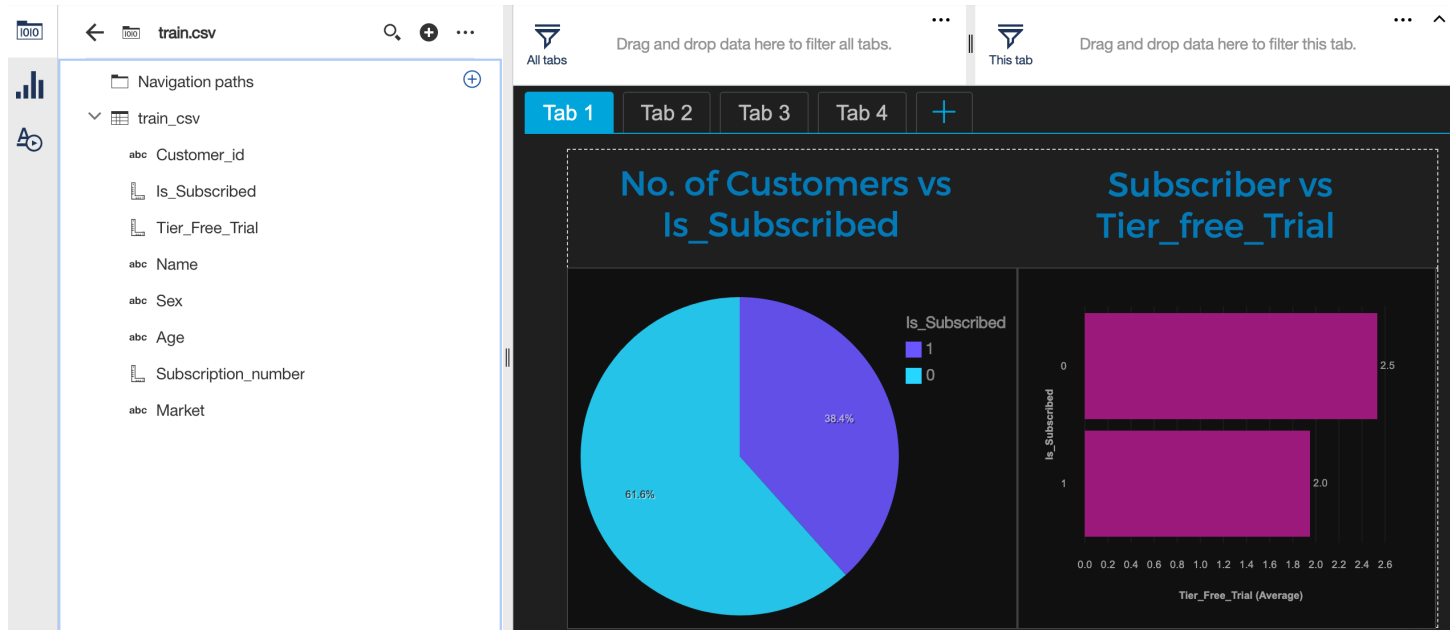
We can see from the above analysis that the subscriber percentage is 38.4% and non subscriber percentage is 61.6%.

Now let's look at the subscriber rate for Tier_Free_Trial.
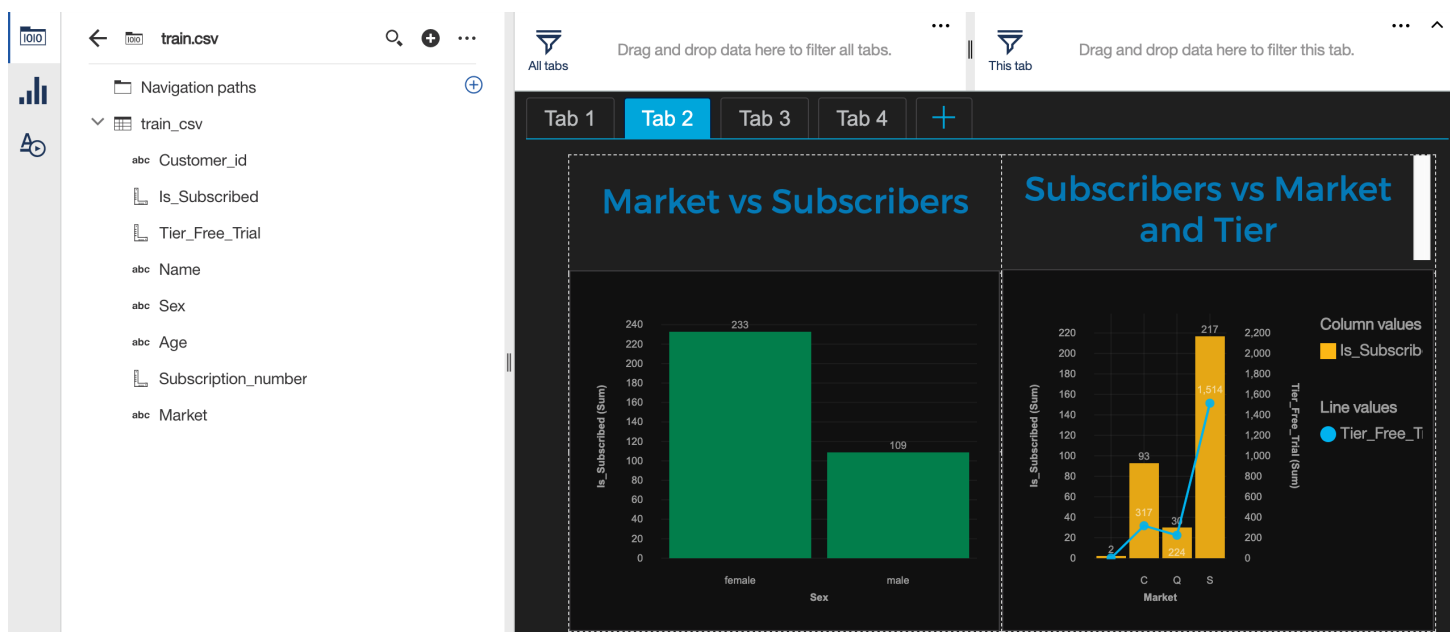
The analysis above shows that 63% customers are from tier 1, 47% from tier 2 and 24% from tier 3 are subscribed out of the total customers. It seems that the customers that embarked from a particular market had a higher rate of subscription at 55%. Let's move forward to see some more data analysis.
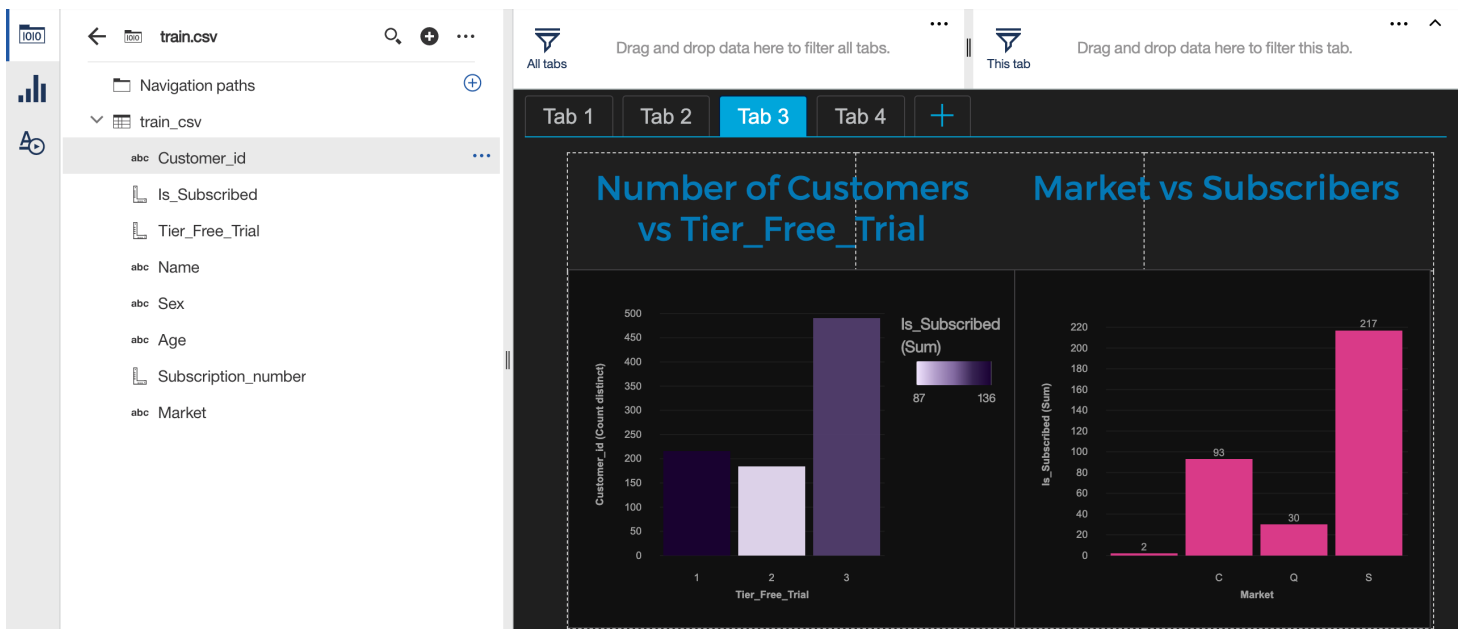
The analysis above shows that the free tier percentage subscribed for a market and the Subscriber_percentage is 35%, 54%, 46%, 25%, 17% respectively. Looks like customers who had either tier 1, 2 or 3 had a higher possibility of subscribing than the ones who had none. However, having more than 3 made the possibility even lesser.
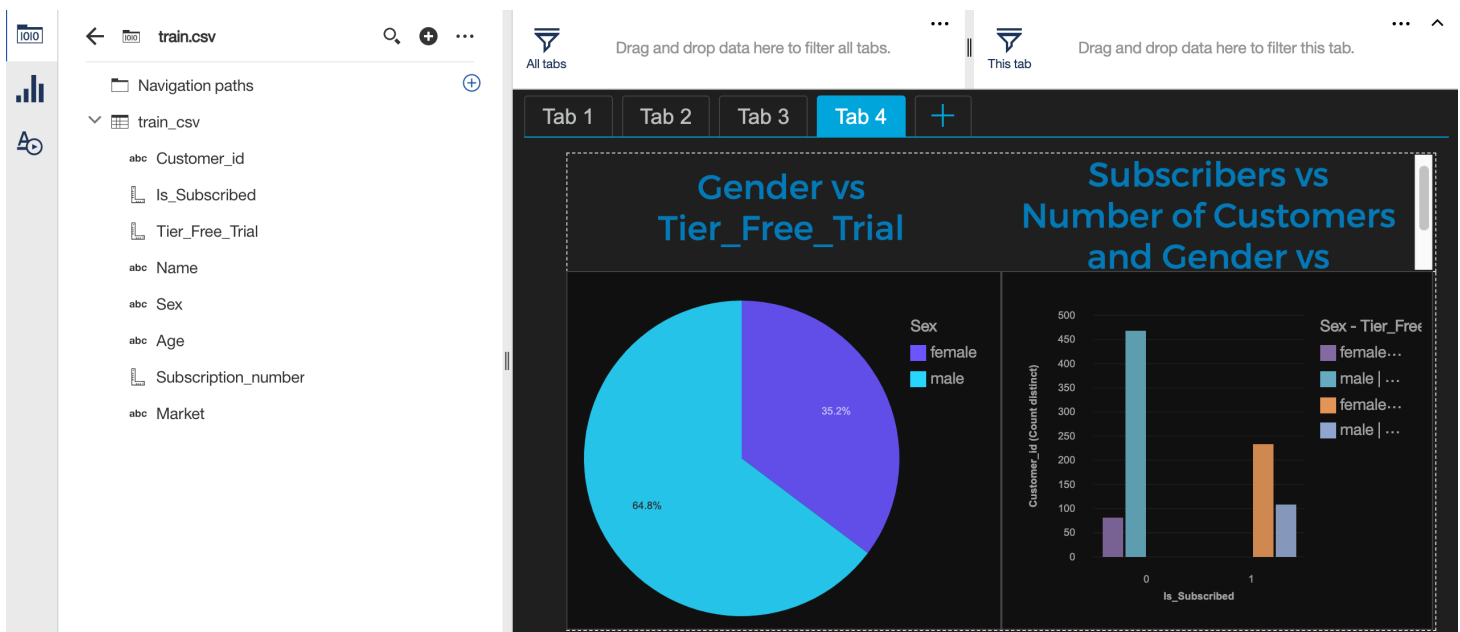
The analysis above shows that from the customers with tier 1, out of 94, 91 female customers subscribed. In the customer tier 2, out of 152 female customers only 70 subscribed and in tier 3 out of 432 female customers only 72 subscribed. For the male customers in tier 1 out of 1,041 only 47 subscribed, in tier 2 out of 122, 45 subscribed and in tier 3 out of 216 only 17 subscribed.

Summary of Findings

The bar plots show the relationship between:

- median subscription, gender and subscriber outcome

- age/gender and subscriber outcome

This simple dashboard demonstrate that features such as subscribed, age and gender can play critical roles in determining the subscriber outcome. Additionally, the features such as Tier_Free_Trial, streams and market can also influence the subscriber outcome. The dashboard from this tutorial can be found here.