**Final-Assignment Report**

This project is designed to simulate a full workflow of a Data Analyst from getting data off the Database to manipulate it with the use of Python and Pandas module to present it through matplotlib module or Tableau.

The concept is that we are given a dataset that contains Liquor Sales in the state of Iowa in USA between 2012-2020 and we are asked to find the most popular item per zipcode and the percentage of sales per store in the period between 2016-2019.

We are also asked to visualize the Data and present them in either a matplotlib format or in Tableau Public.

Every calculation and transformation of Data has to happen through a Python Script.

###### Step 1.

Add the Dataset provided to Workbench.

###### Step 2.

Use a Query to get all the columns of the table between the years 2016-2019. The query we write to display the dates is “SELECT \* FROM `finance\_liquor\_sales` where `date` between '2016-1-1' and '2019-12-31' ;”

###### Step 3.

Export the data to an CSV file, creative a new project in PyCharm and save the CSV file.

###### Step 4.

Use Python and Pandas to Aggregate the CSV data so we can get the most popular item sold based on zip code.

import matplotlib.pyplot as plt  
import pandas as pd  
import numpy as np  
  
df = pd.read\_csv("finance\_liquor\_sales.csv")  
cn = df.groupby(['zip\_code','bottles\_sold'])  
print(cn.first().to\_string())

###### Step 5.

Use Matplotlib with the newly made CSV file and present your Data.

plt.plot([50000, 53000], [0, 1000])  
plt.title("Bottles Sold")  
plt.xlabel("X - Zip Code")  
plt.ylabel("y - Bottles Sold")  
df.plot(kind='scatter',x='x',y='y')  
plt.show()