**Project**

**Value Shopper**

**Submitted by**

**Problem Statement -** We are building an E-commerce enterprise data platform in Cloud for Customer ‘Slalom’ to consolidate their segregated platforms into one environment in cloud that reduces their operations cost and helps perform an efficient analytics which they are unable to perform today.

1. Technical Rationalization - Which platform, technologies, tools, services and language would you choose and why?

* My main focus was on using Open source tools as the solution should also be cost effective.
* Data Extraction, Exploratory Data Analysis, Merging the files was done using Python to create a Python dataframe which was later extracted as an csv file.
* Used Google Colab as the file size was too high and required high processing speed which Google Colab provides free of cost.
* Later the data set was consumed by Power BI one of the Analytics tool to draw insights.
* Final deliverable was the report published on Power BI workspace which can be shared and reviewed by peers.

4.Architecture diagram describing the services you have identified and how you envision an end to end operational platform.

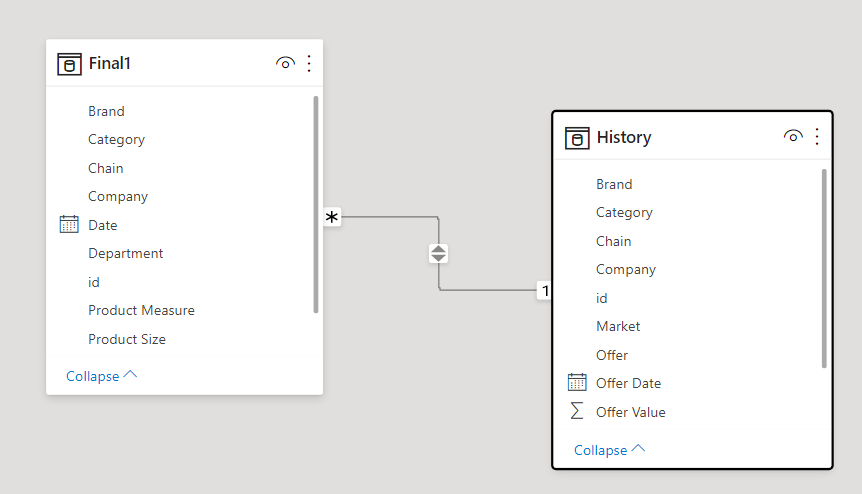
* Load Offer Table and Train History table and further merge it on ‘Offer’ and create a new table History.

|  |
| --- |
| **TrainHistory** |
| id |
| chain |
| Offer |
| market |
| repeattrips |
| repeater |
| offerdate |

|  |
| --- |
| **Offer Table** |
| Offer |
| Category |
| quantity |
| company |
| offervalue |
| brand |

|  |
| --- |
| **History** |
| id |
| chain |
| offer |
| market |
| repeattrips |
| repeater |
| offerdate |
| Category |
| quantity |
| company |
| offervalue |
| brand |

* Join History table with Transaction Table (Final1 Table) in Power BI on ID column.

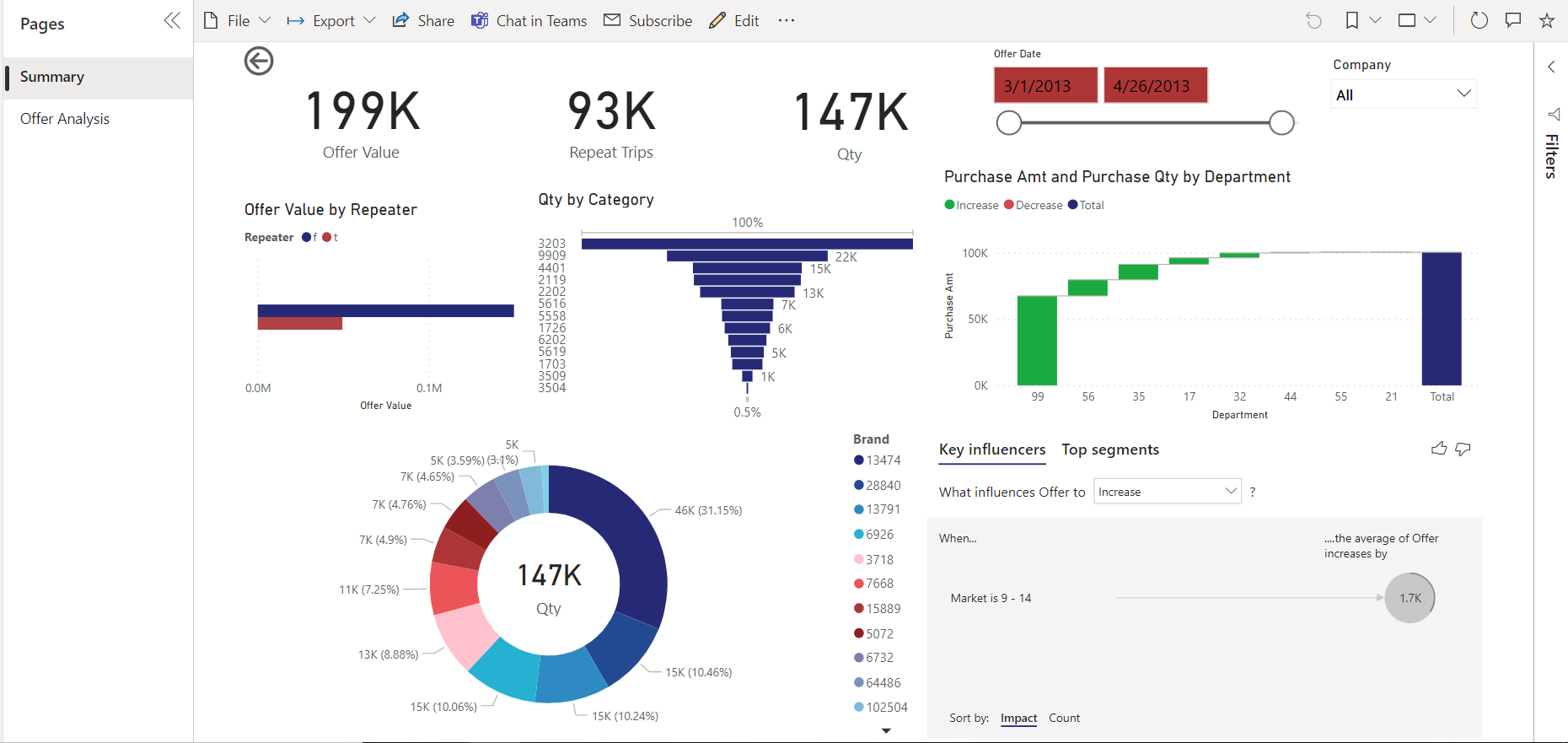


Note: Transaction table has data before for before offer date and after date, considering the file size Transaction table was filtered on offer date and it was divided into 18 different chunks and chuck\_0 which we call Final1 was taken for analysis.

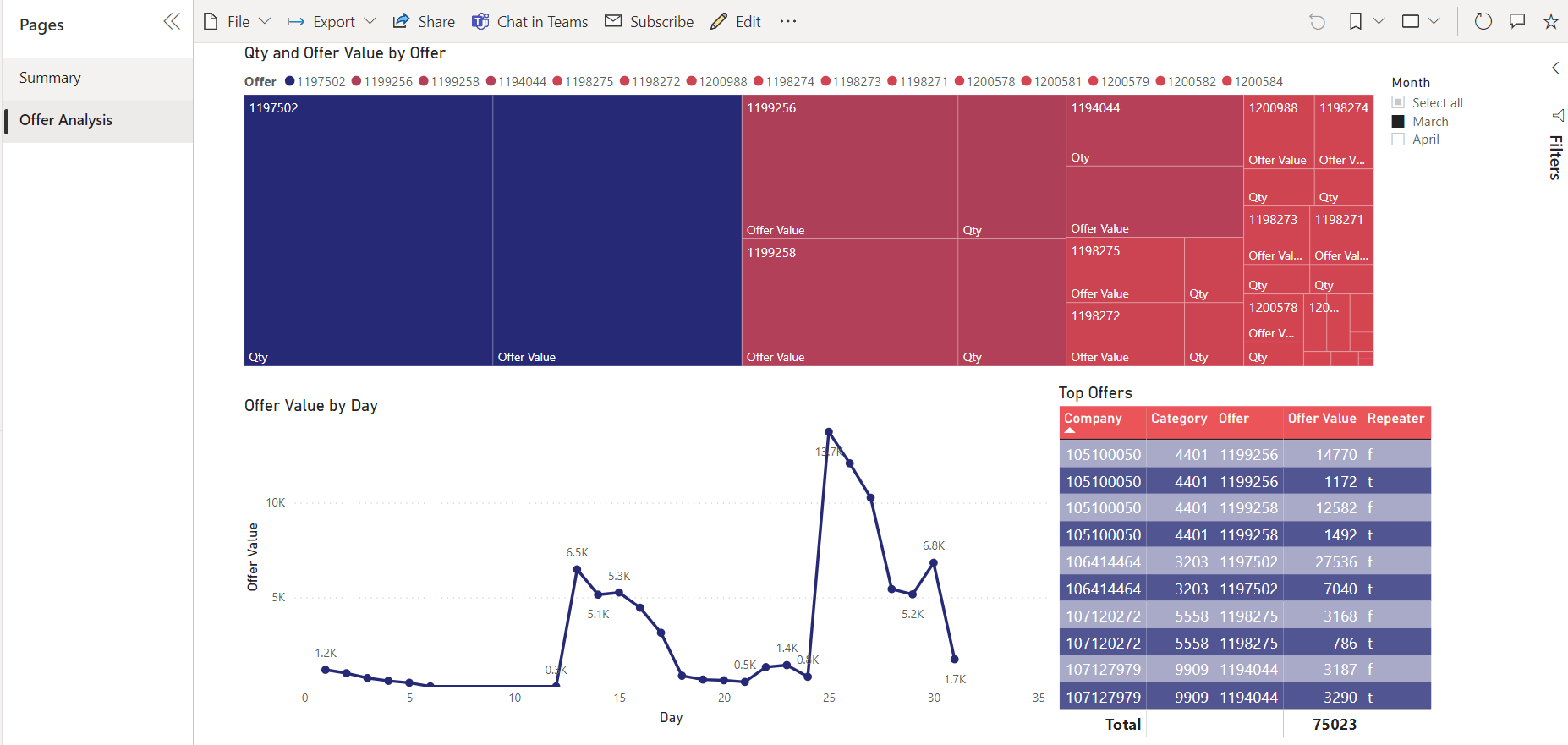
5.Pease visualize the data by creating an application within Tableau,

QlikView or QlikSense, or Power BI. Feel free to highlight your skills by using the highest level of sophistication

* Summary Page



* Offer Analysis Page



6.Describe approach towards productionization of your solution

Productionization is a design objective and good robust design is a key.

Productionization involves up-front investment in systems that smooth the deployment, maintenance, and adoption of whatever data processes we choose to employ. The design work necessary for productionization almost always lengthens the time it takes to launch a product, and because of that it is often neglected. But a delayed launch is less frustrating and expensive than a blundered launch. It pays to take the time to design the system well.

Fully productionized data science is a circle, not a line.

If we do that, then instead specifying then conducting and then evaluating the analysis, we have just two tasks — operation and calibration — and each task feed into the other. The point of measurement shouldn’t be to spin out a report, but rather to adjust the initial parameters of the analysis while that analysis is still in progress (the report can be produced as a side-effect of the calibration effort). The initial parameters should be treated as priors to be modified rather than as requirements to be met.

Best approach as follows [business needs] -> [algorithms] -> [results]. It is preferable to think about it as [business needs] -> [algorithms] -> [revised needs] -> [revised algorithms] -> [revised needs], and so on.

**THANK YOU**