**Team -03**

Amazon

Amazon:

Amazon.com, Inc., is an American multinational technology company based in Seattle, Washington that focuses on e-commerce, cloud computing, digital streaming and artificial intelligence. It is considered one of the Big Four technology companies along with Google, Apple and Facebook.

Amazon is known for its disruption of well-established industries through technological innovation and mass scale. It is the world's largest e-commerce marketplace, AI assistant provider, and cloud computing platform as measured by revenue and market capitalization. Amazon is the largest Internet company by revenue in the world. It is the second largest employer in the United States and one of the world's most valuable companies. Amazon is the second largest technology company by revenue.

Amazon was founded by Jeff Bezos on July 5, 1994 in Bellevue, Washington. The company initially started as an online marketplace for books but later expanded to sell electronics, software, video games, apparel, furniture, food, toys, and jewelry. In 2015, Amazon surpassed Walmart as the most valuable retailer in the United States by market capitalization. In 2017, Amazon acquired Whole Foods Market for $13.4 billion, which vastly increased Amazon's presence as a brick-and-mortar retailer. In 2018, Bezos announced that its two-day delivery service, Amazon Prime, had surpassed 100 million subscribers worldwide.

Amazon distributes downloads and streaming of video, music, audiobook through its Amazon Prime Video, Amazon Music, and Audible subsidiaries. Amazon also has a publishing arm, Amazon Publishing, a film and television studio, Amazon Studios and a cloud computing subsidiary, Amazon Web Services. It produces consumer electronics including Kindle e-readers, Fire tablets, Fire TV, and Echo devices. In addition, Amazon subsidiaries also include Ring, Twitch.tv, Whole Foods Market, and iMDb.

Dataset:

This is a list of over 28,000 consumer reviews for Amazon products like the Kindle, Fire TV Stick, and more provided by Datafiniti's Product Database. The dataset includes basic product information, rating, review text, and more for each product.

Our team attempted to clean, process and analyze the data and separated it into B2B and B2C relations. We also separated the fields representing Products, Orders, Manufacturers, Customers details.

Dataset:

Customer:This table has details about customers like their username, reviwes, orders, etc.

Manufacturer: This table contains details about Amazon product manufacturer’s.

Product: This table contains details about Amazon products, their price, category of product etc.

Promotions: This table contains information about the various promotions that Amazon had during the year of 2016 and 2017 to increase their sales of their products.

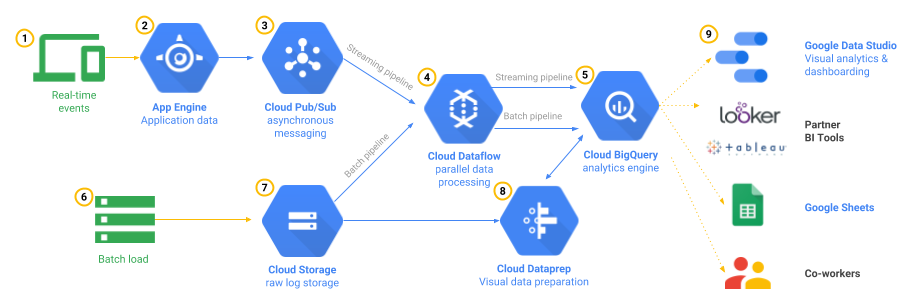
**What can be done with this dataset:**

You can use this data to analyze Amazon’s most successful consumer electronics product launches; discover insights into consumer reviews and assist with machine learning models. E.g.:

* What are the most reviewed Amazon products?
* Tracking Unique Purchases (Unique Purchases refers to the total number of times a product (or set of products) was part of a transaction.)
* Calculation the Average order Value in order to understand customer spending habits.
* What are the initial and current number of customer reviews for each product?
* How do the reviews in the first 90 days after a product launch compare to the price of the product?
* How do the reviews in the first 90 days after a product launch compare to the days available for sale?
* Tracking future sales based on the quantity (units sold).

Marketing Analytics System:

We have setup Google Cloud Platform (GCP) as the system to build the basic pipeline.



The above image depicts the GCP pipeline we will be implementing.

**App Engine**:

It is a fully managed, serverless platform for developing and hosting web applications at scale.

**Cloud Pub/Sub**:

It is a fully-managed real-time messaging service that allows you to send and receive messages between independent applications.

**Cloud Dataflow**:

Cloud Dataflow is a managed service for executing a wide variety of data processing patterns.

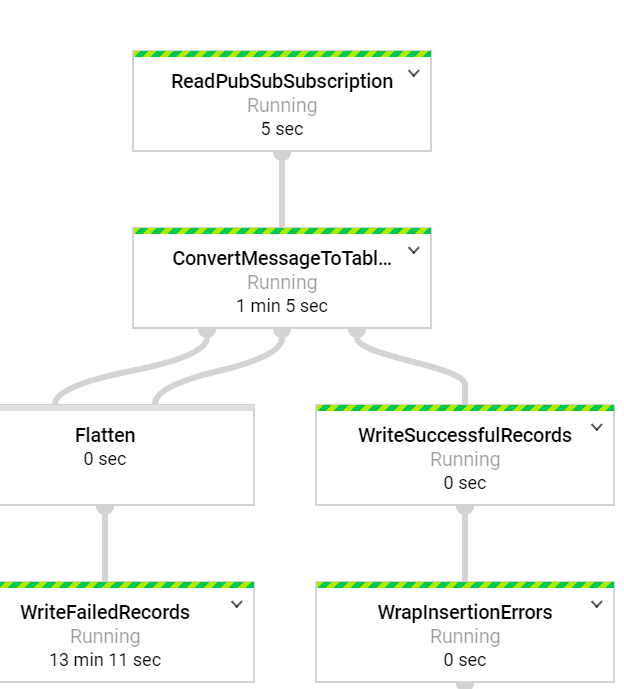
**Cloud Big query**:

Big Query is Google's data warehouse Analyze all your batch and streaming data by creating a logical data warehouse over managed columnar storage, as well as data from object storage and spreadsheets.

We can process real time data as well as batch time events to the GCP pipeline architecture.

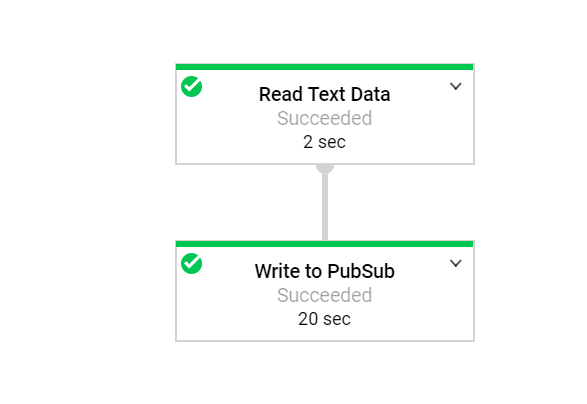
In case of **real time events** being triggered, we will be following steps:

1. Cloud Pub/Sub ingests and temporarily stores these messages.
2. Cloud Dataflow transforms the JSON event into structured, schema-based data.
3. That data is loaded into the Big Query analytics engine.



In case of **batch time events**, we can perform direct loading from Cloud Storage to Big Query.

We can directly load the files from CSV files in Cloud Storage by providing a schema and starting a load job



Steps to Conduct GCP:

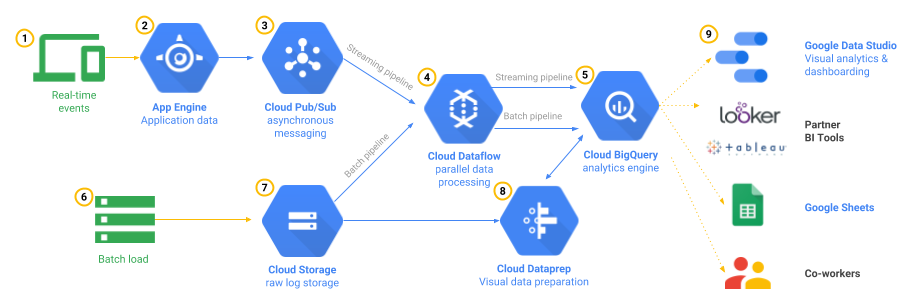
1. Create a GCP Account using an E-mail account with free Student Credits of $300

<https://cloud.google.com/blog/products/gcp/new-google-cloud-platform-education-grants-offer-free-credits-to-students>

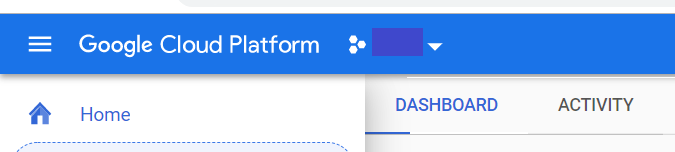
1. Mobile Gaming Analytics Pipeline Documentation:

Read: <https://cloud.google.com/solutions/mobile/mobile-gaming-analysis-telemetry>

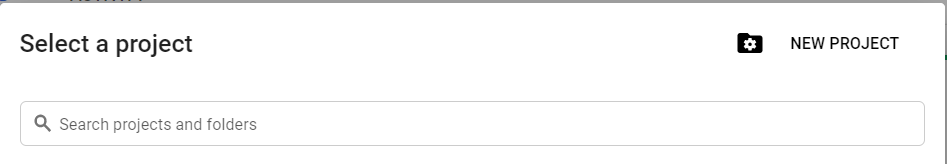
1. Architecture Under Consideration:



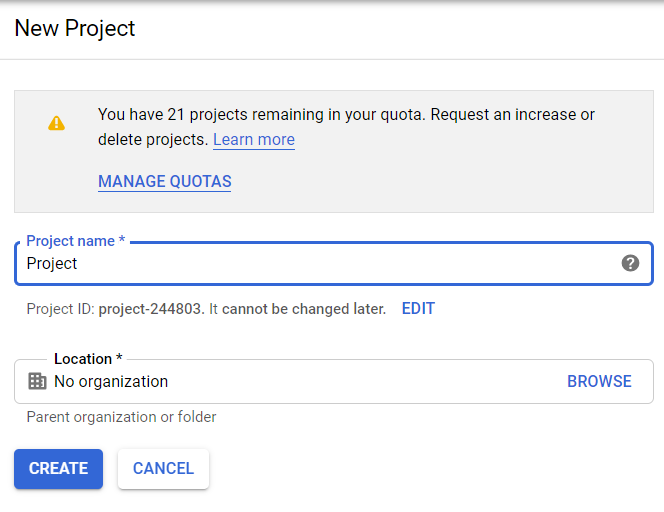
1. Create a Project
2. Login to GCP Console
3. Select Top Left Drop down:

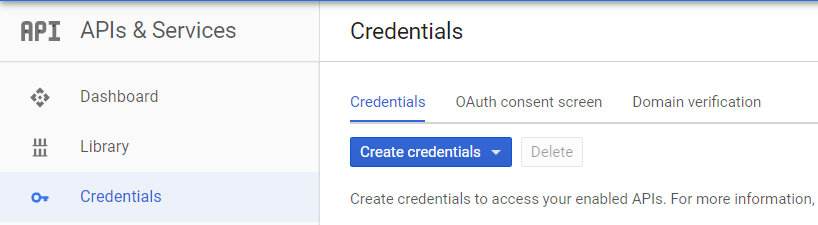


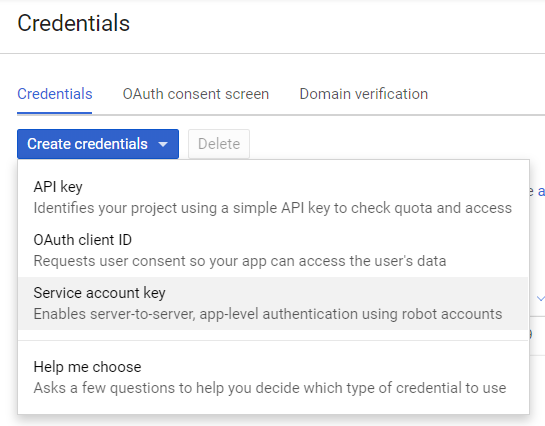
1. Click on New Project



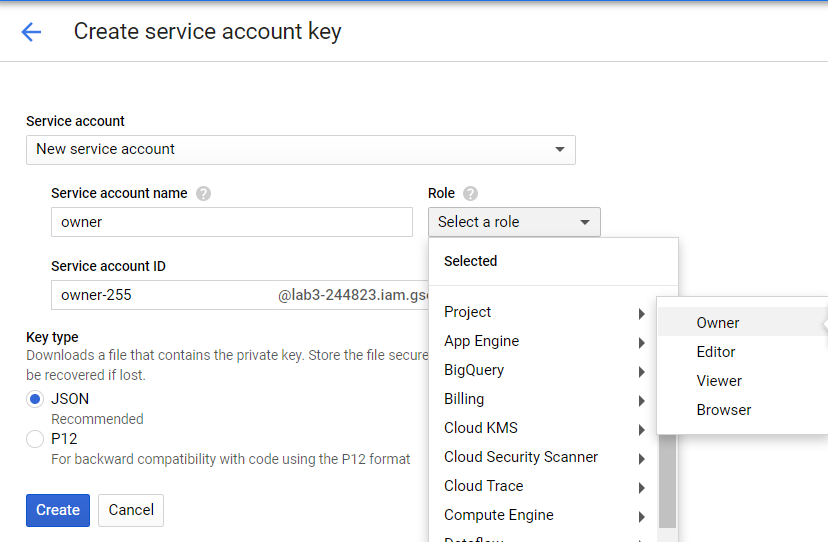
1. Create New Project



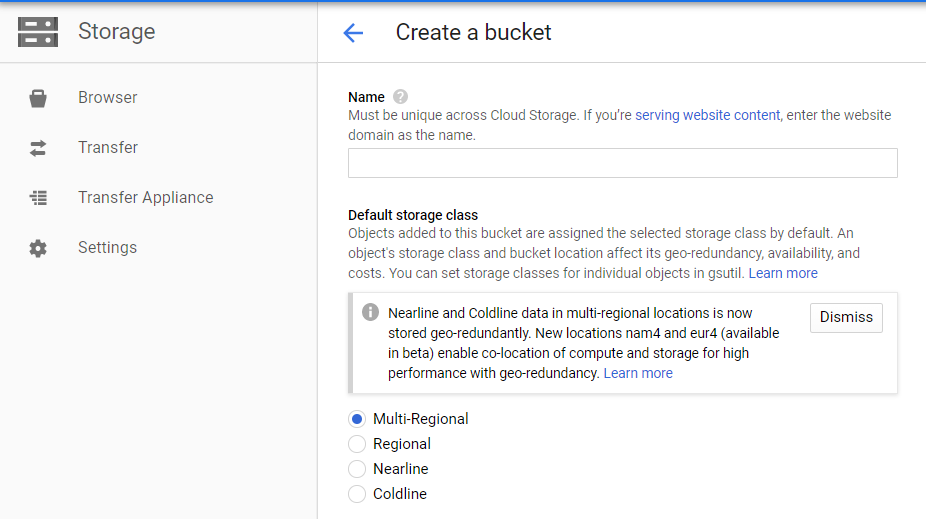
1. Create a credential file to access the Project from API.
2. Click on: API & Services from GCP console. 
3. Click on Create credentials of type Service account Key:



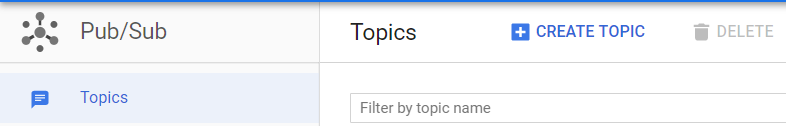
1. Create a service account name with Project Owner Role for your Project.

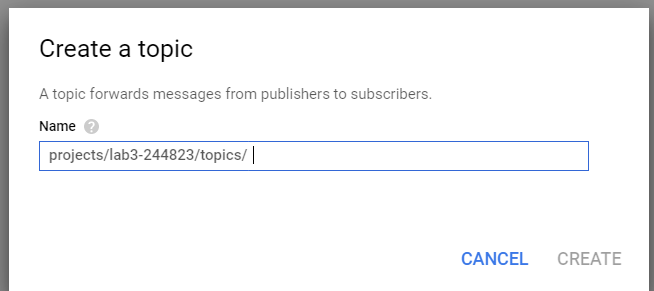


1. Create a Cloud Storage Bucket.

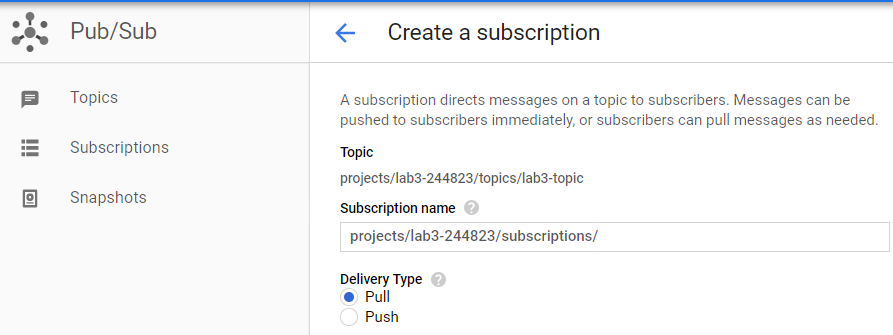


1. Create a PubSub topic





1. Create a PubSub Subscription for the topic created.

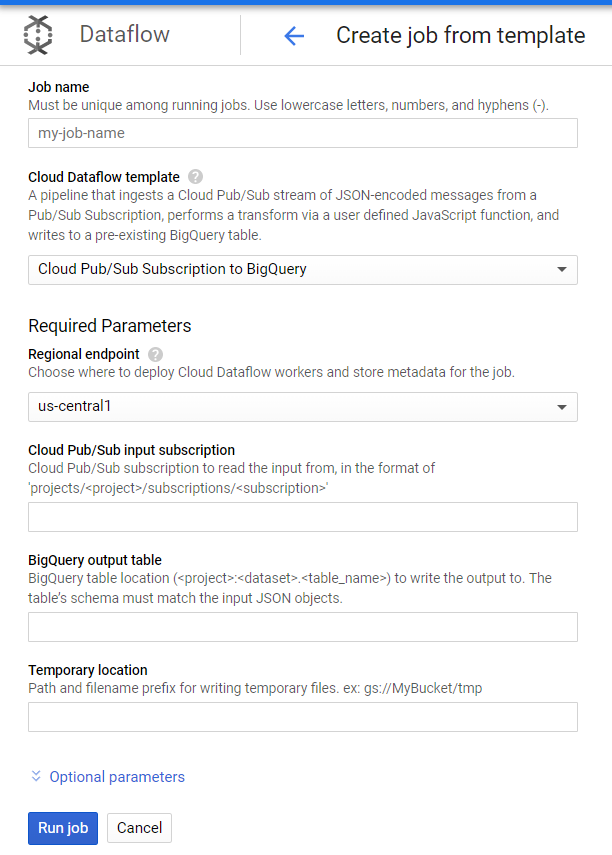


1. Create a Bigquery Table:

a. Create a BigQuery Dataset.

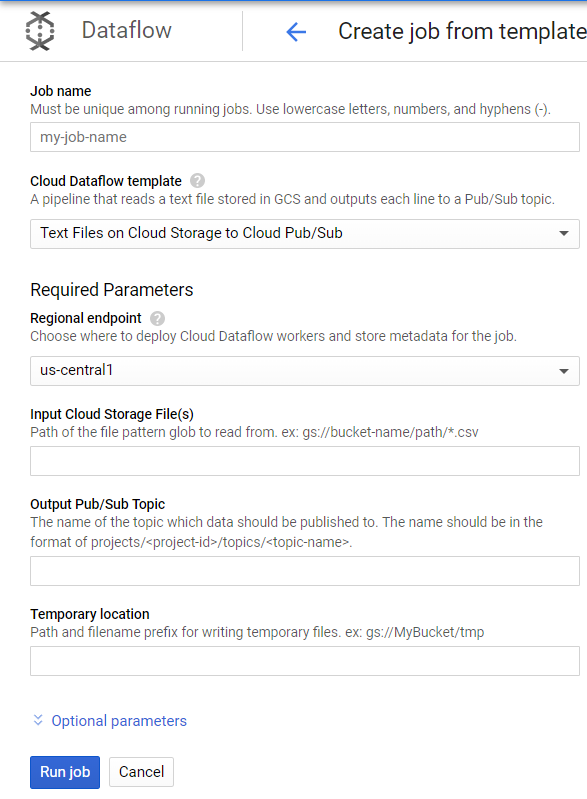
b. Create a BigQuery Table with the required schema under the created dataset.

10. Create a Streaming DataFlow for PubSub Subscription to Bigquery Table.

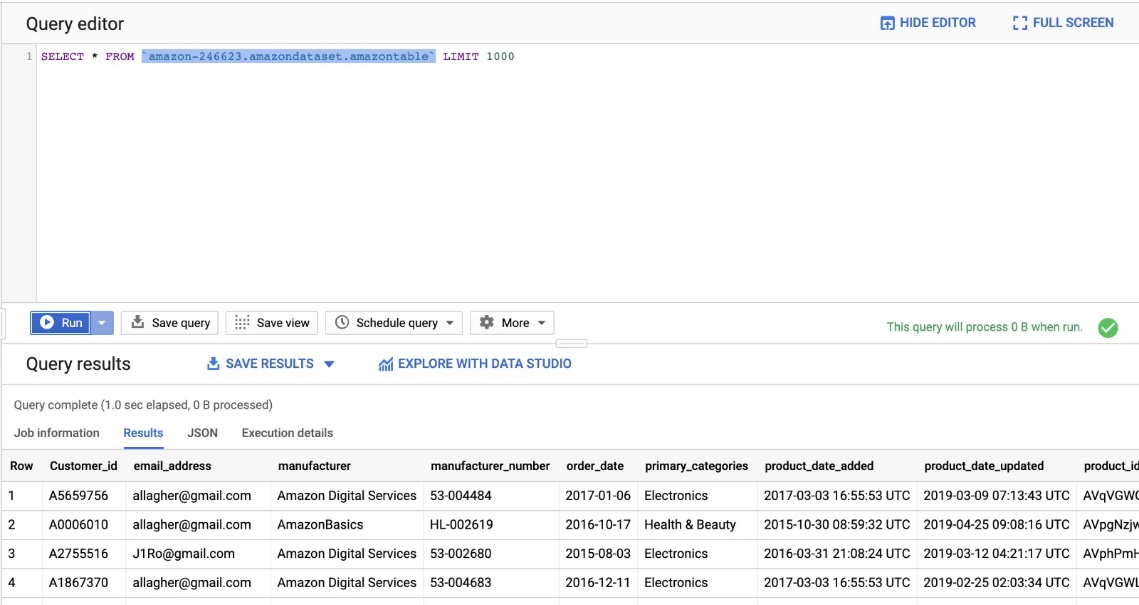


1. Creating a Batch Processing Dataflow.

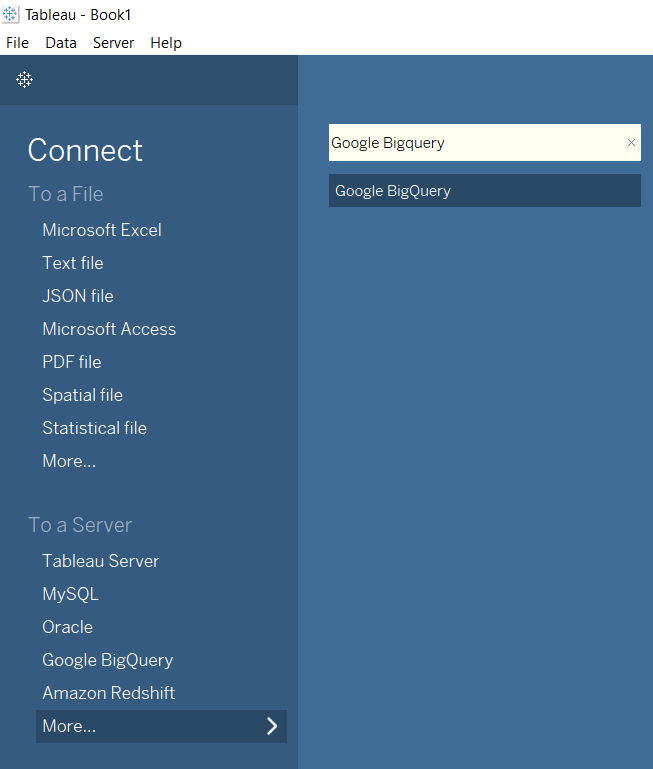
Give the cloud storage path of the data file.



1. Go to BiqQuery Table and query the table for the Data load.



1. Connecting BigQuery Table to Tableau



Amazon Problem Statement:

We faced some challenges in finding a problem statement with Amazon as it has several departments and categories of products and is doing well among its competitor’s. So, we decided to investigate the dataset that we had and tried to find problems with respect to customers and sales of products and performed some analytics.

One problem that we found with our data is with respect to business to customers (B2C) relations where the sales of amazon products is good in some regions of United States but is poor in other regions.

Another problem that found with respect to business to business(B2B) relations is that the sales of certain products where not that consistent. For instance, customer’s bought electronics products mainly during certain times of the year and not during other months. This information helps us understand customer buying patterns, Uniqueness in purchases for different products.

With the given problem statements, we are going to come up with possible solutions in order to have gateway to higher conversion in Region where the sales are currently not doing well, as well how to increase the purchase rate.

Access to our system:

Giving GCP BigQuery Access to another team:

1. Login to GCP Cloud Console

2. Select your working Project

3. Go to IAM service

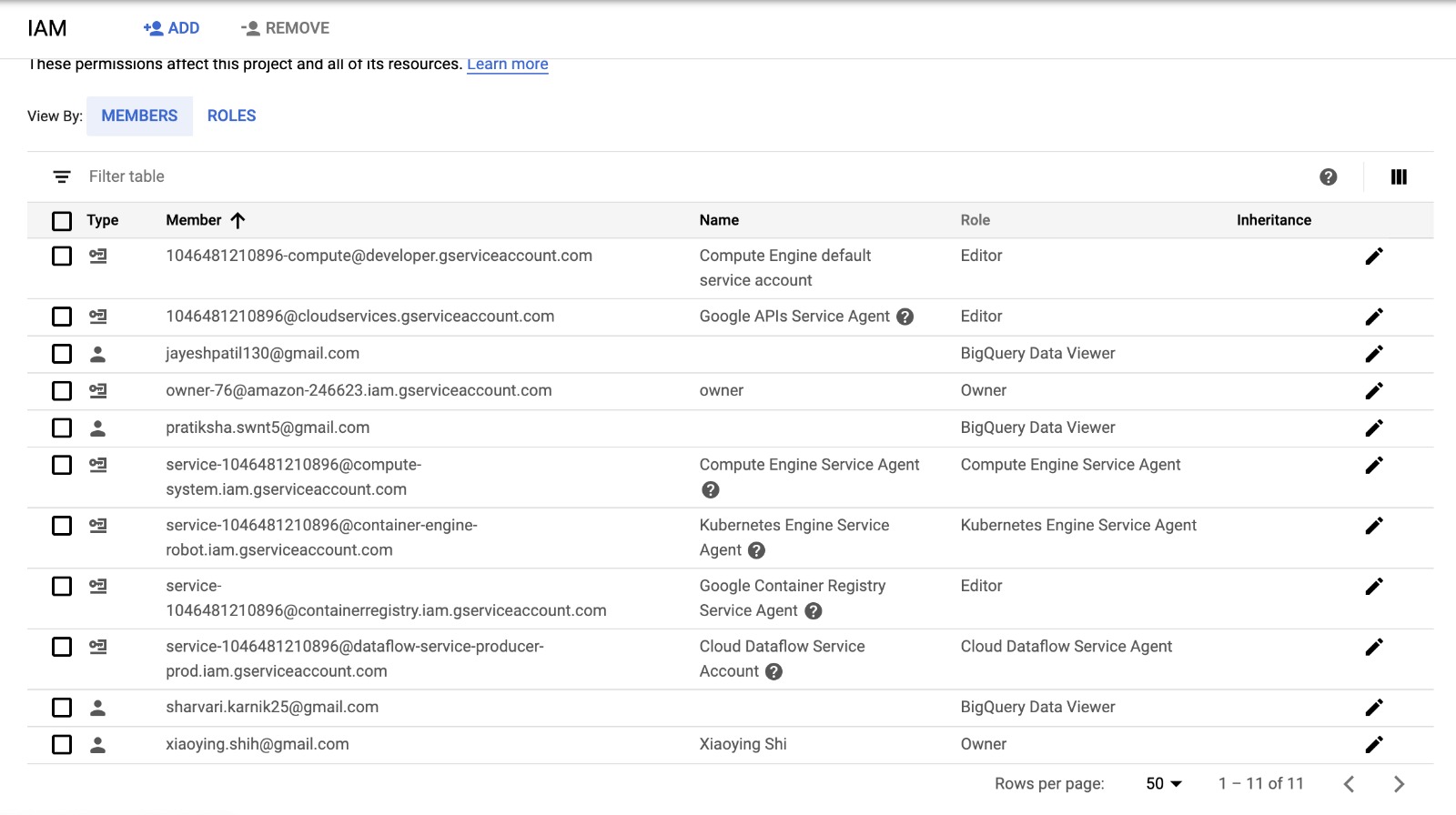
4. Click on +Add member

5. Give the email id of user in the New member section

6. Select Role as BigQuery: BigQuery Data Viewer

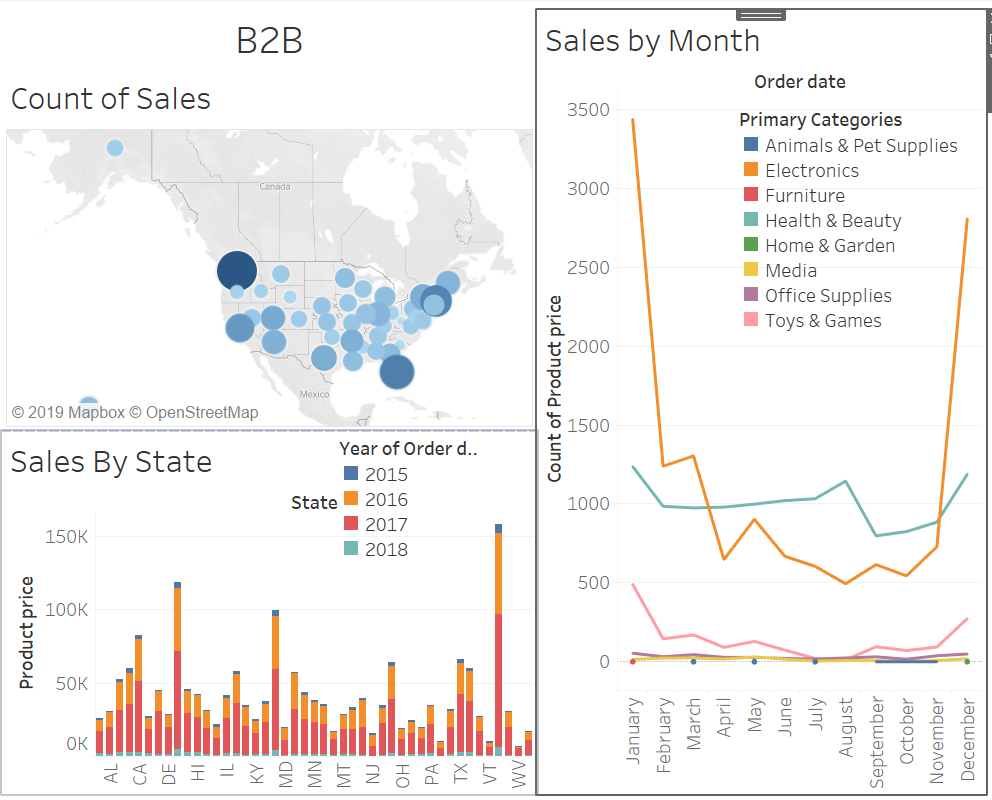
7. Save

8. The added member now able to see the Dataset and Table in the Tableau connector

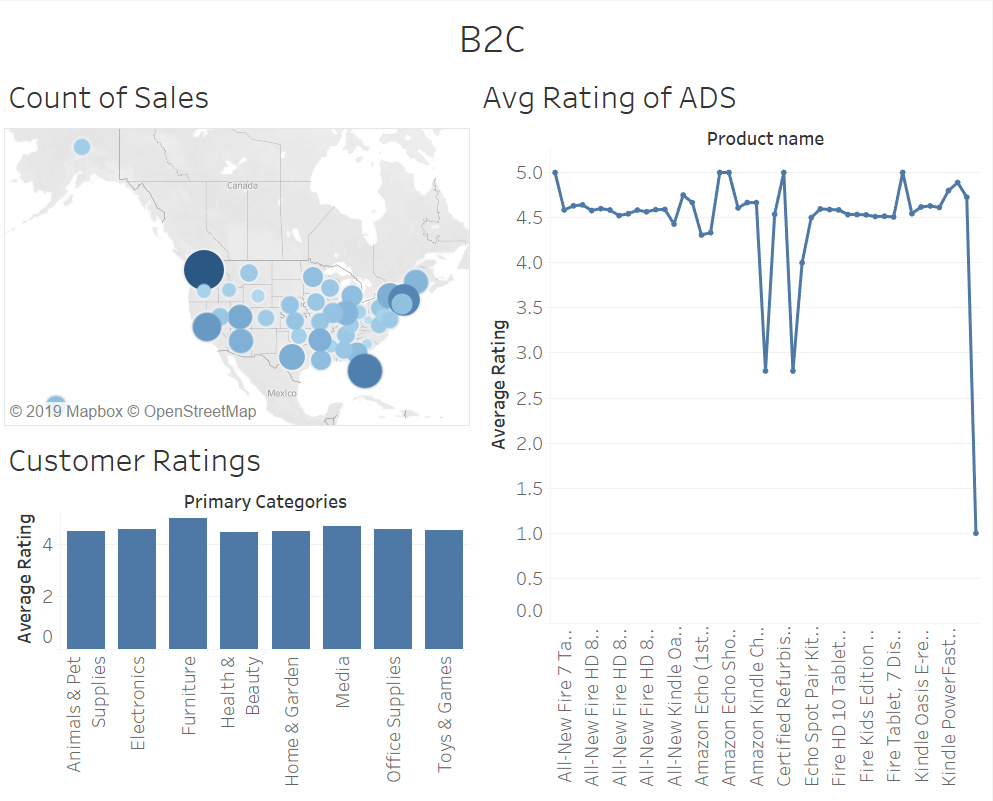


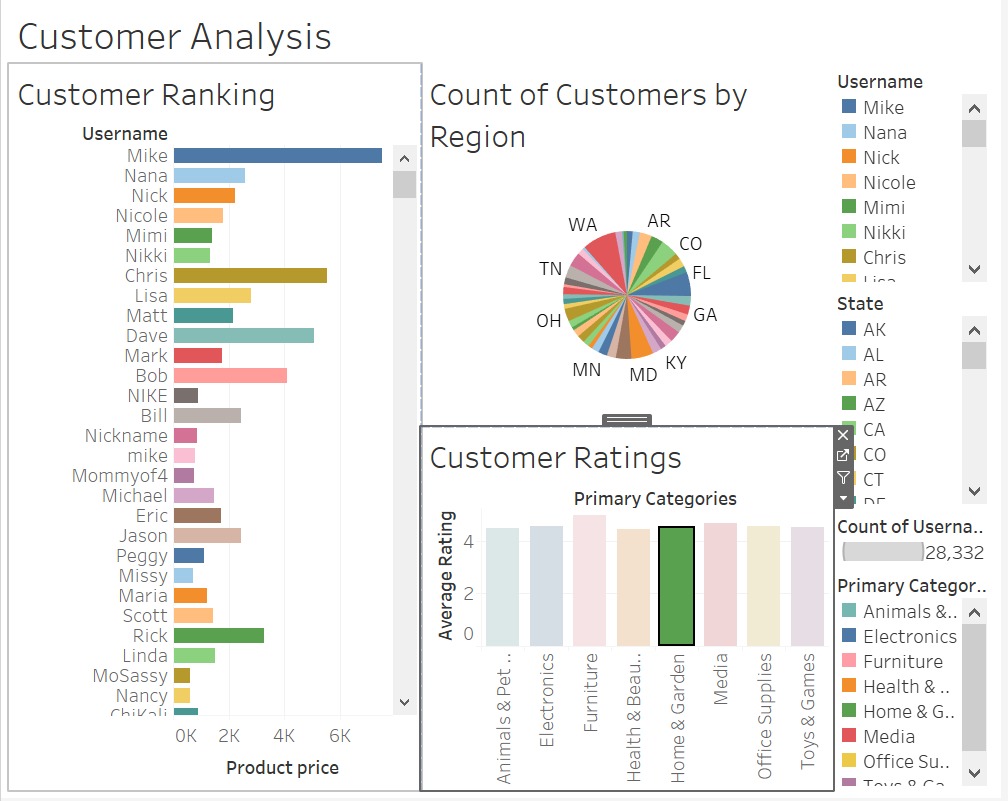
Analytics Dashboard:

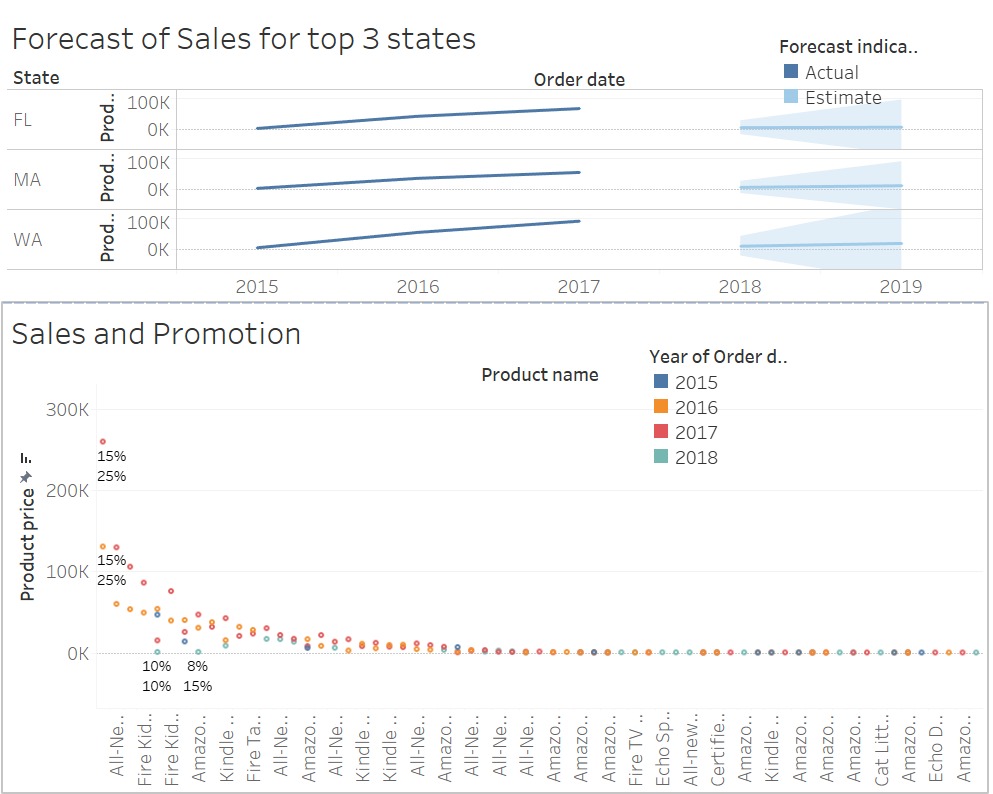
B2B:



B2C:







INSIGHTS:

SALE DASHBOARD:

Count of Sales and Sales by state :

From the above graph we wanted to check the sales per region and it shows that sales in all the region is not consistent. Some regions like Washington, Florida show high sales while regions like West Virginia show poor sales.

Sales by Month:

We tried to get patterns as to when the customers are buying our products for each category. We can see that for the category Electronics customer tend to buy in January, November and September which have big sale days like Prime Day, Thanksgiving and Labour day respectively. Customer tend to buy electronics on the big sale days only and rest of the year the sales stay contant.

Also for other categories, Like Toys and Furniture, there are lot of competitors, hence the sales needs to be fixed with some promotional discounts/offers.

Goal:

Our main goal is to improve the sales all over USA and have consistent and profitable sales in each category. We mainly focused on the categories which had a decent sale in 2015 and can make better sales in 2016 and 2017 by adding promotion/offers/discounts.

Solution:

We added promotion for categories to have better sales in 2016 and 2017. And from the graph we can see that growth is very good after the discounts when compared to 2015 sales. Though few category sales were still lagging behind to find how the sales for these categories In the future

Future Solutions:

For More better sales in the coming future, we can concentrate on the purchasing patterns of customer and make sure to retain the customers who have the highest purchasing rate by engaging with through emails, or any other social referral sites or give them offers, giftcards.

Also we can segment the customers into different personas like corporate, consumer etc and build a target market accordingly. We can also  track Each persona as they will have buyer journey already understood to some degree; what you can do now is plan how you intend to support that journey at each stage.

To improve the Consistency of sales in every Category, we can compare prices with the competitors and make sure we have a competitive price. Also by check our market positioning and what exactly our audience is talking about our products and services

References:

* About Amazon:

<https://en.wikipedia.org/wiki/Amazon_(company)>

* Dataset:

<https://data.world/datafiniti/consumer-reviews-of-amazon-products/workspace/project-summary>

* Marketing Analytics System Architecture:

<https://cloud.google.com/solutions/mobile/mobile-gaming-analysis-telemetry>

* Getting Access to our System Data:

<https://console.cloud.google.com/iam-admin/iam?project=assignment69216&folder=&organizationId=>

* Tableau:

<https://onlinehelp.tableau.com/current/pro/desktop/en-gb/gettingstarted_overview.htm>

Team Members:

Anisha Ganguly

Xiaoying Shi

Ruta Lad