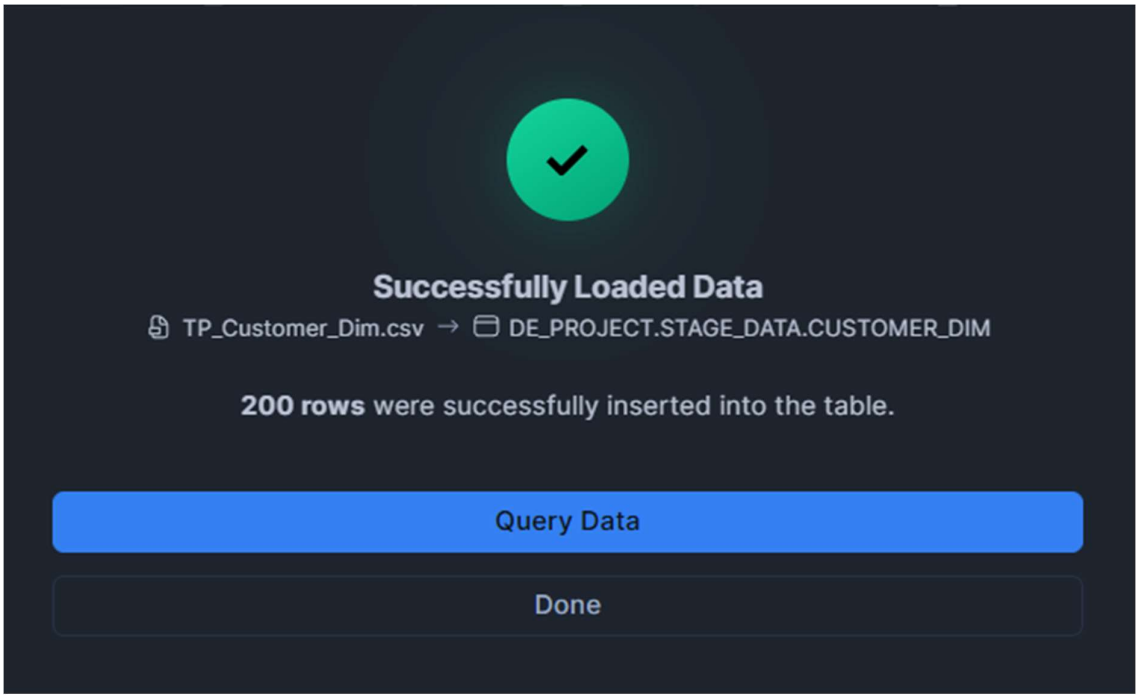
Project Deliverable Part 3

To reflect upon Part 2, where I have created the staging and production schemas and tables using Snowflake to upload the data for cleaning by removing the null values, and duplicates, transforming certain fields, and formatting it for data integrity and consistency.



Results Chart

	ORDER_ID	ORDER_AMT	ORDER_QTY	CUSTOMER_ID	ORDER_DATE	NEW_ORDER_DATE	NEW_ORDER_TIME
1	1	543.50	3	20	2022-07-14 15:43:22.000 +0000	2022-07-14	15:43:22
2	2	138.62	4	156	2022-09-25 08:25:31.000 +0000	2022-09-25	08:25:31
3	3	421.68	1	6	2023-07-22 02:54:07.000 +0000	2023-07-22	02:54:07
4	4	197.81	3	153	2022-03-20 00:00:00.000 -0700	2022-03-20	00:00:00
5	5	742.61	3	89	2023-10-05 17:16:24.000 +0000	2023-10-05	17:16:24
6	6	832.69	1	135	2022-03-10 22:21:33.000 +0000	2022-03-10	22:21:33
7	7	363.86	5	54	2023-04-19 01:44:14.000 +0000	2023-04-19	01:44:14
8	8	371.74	3	125	2022-04-19 04:37:36.000 +0000	2022-04-19	04:37:36
9	9	784.62	1	196	2023-02-01 15:28:41.000 +0000	2023-02-01	15:28:41
10	10	631.18	1	167	2022-12-04 13:41:21.000 +0000	2022-12-04	13:41:21
11	11	79.39	4	96	2022-02-16 08:14:18.000 +0000	2022-02-16	08:14:18

Finally, I have uploaded the cleaned and transformed data into the production table.

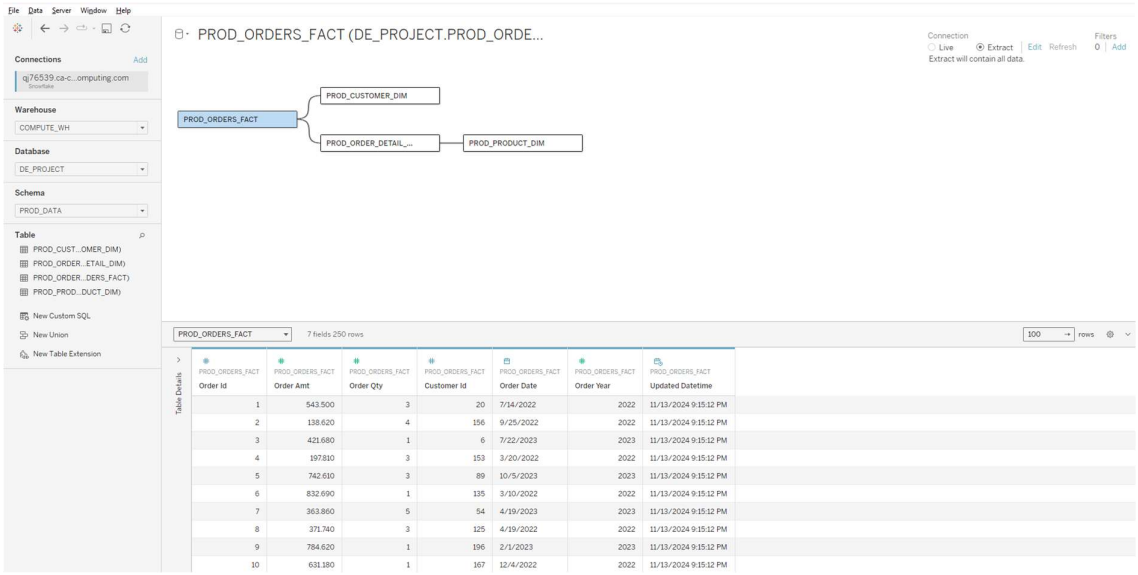
```
198
199 SELECT * FROM DE_PROJECT.PROD_DATA.PROD_CUSTOMER_DIM;
200
201
202
203
204
205
```

Results Chart

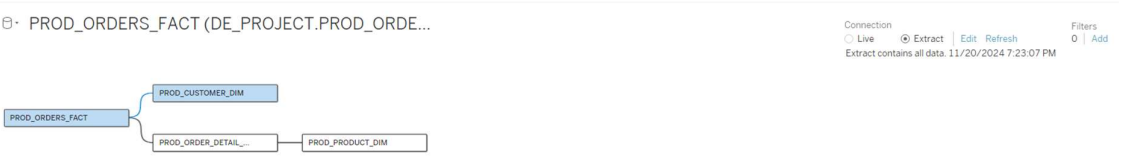
	CUSTOMER_ID	FIRST_NAME	LAST_NAME	EMAIL	PHONE	CUST_ADDRESS	CUST_STATE	CUST_CITY	CUST_ZIP	UPDATED_DATE
1	1	Penelope	Keddie	pkeddie0@behance.net	205-151-7125	3 8th Crossing	Alabama	Birmingham	35244	2024-11-13 18:02:24.717 -0800
2	3	Jake	Gorger	jgorger2@springer.com	208-562-9998	68153 Maryland Center	Idaho	Boise	83757	2024-11-13 18:02:24.717 -0800
3	5	Angela	Nattriss	anattriss4@shinystat.com	513-792-4285	02 Maryland Hill	Ohio	Cincinnati	45249	2024-11-13 18:02:24.717 -0800
4	7	Lyman	Gascoyen	lgascoyen6@admin.ch	214-402-0630	045 Gale Street	TX	Dallas	75277	2024-11-13 18:02:24.717 -0800
5	8	Carlina	Shilliday	cshilliday7@shareasale.com	814-114-4016	708 Memorial Hill	Pennsylvania	Erie	16510	2024-11-13 18:02:24.717 -0800
6	9	Maribel	Guile	mguile8@techcrunch.com	808-103-4647	5421 Myrtle Lane	Hawaii	Honolulu	96820	2024-11-13 18:02:24.717 -0800
7	10	Della	Pisculli	dpisculli9@slashdot.org	530-256-6967	41 Jackson Pass	CA	South Lake Tahoe	96154	2024-11-13 18:02:24.717 -0800
8	11	Tatiana	Bonner	tbonnera@php.net	719-966-7996	99687 Hagan Road	Colorado	Colorado Springs	80920	2024-11-13 18:02:24.717 -0800
9	12	Nobe	Folli	nfollib@columbia.edu	405-131-6665	5 Roth Way	Oklahoma	Oklahoma City	73135	2024-11-13 18:02:24.717 -0800

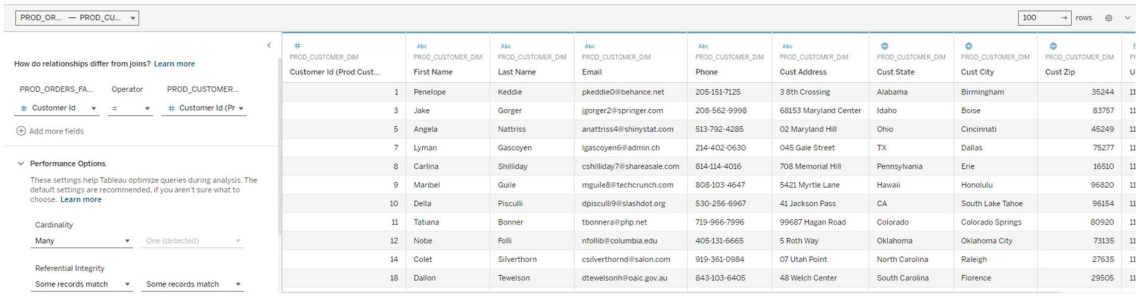
Now, I'm going to utilize the data that I have ingested, transformed, and loaded data into our prod schema earlier before by establishing a connection between snowflake and tableau desktop to do analysis and create dynamic visualizations to extract meaningful insights from our transformed

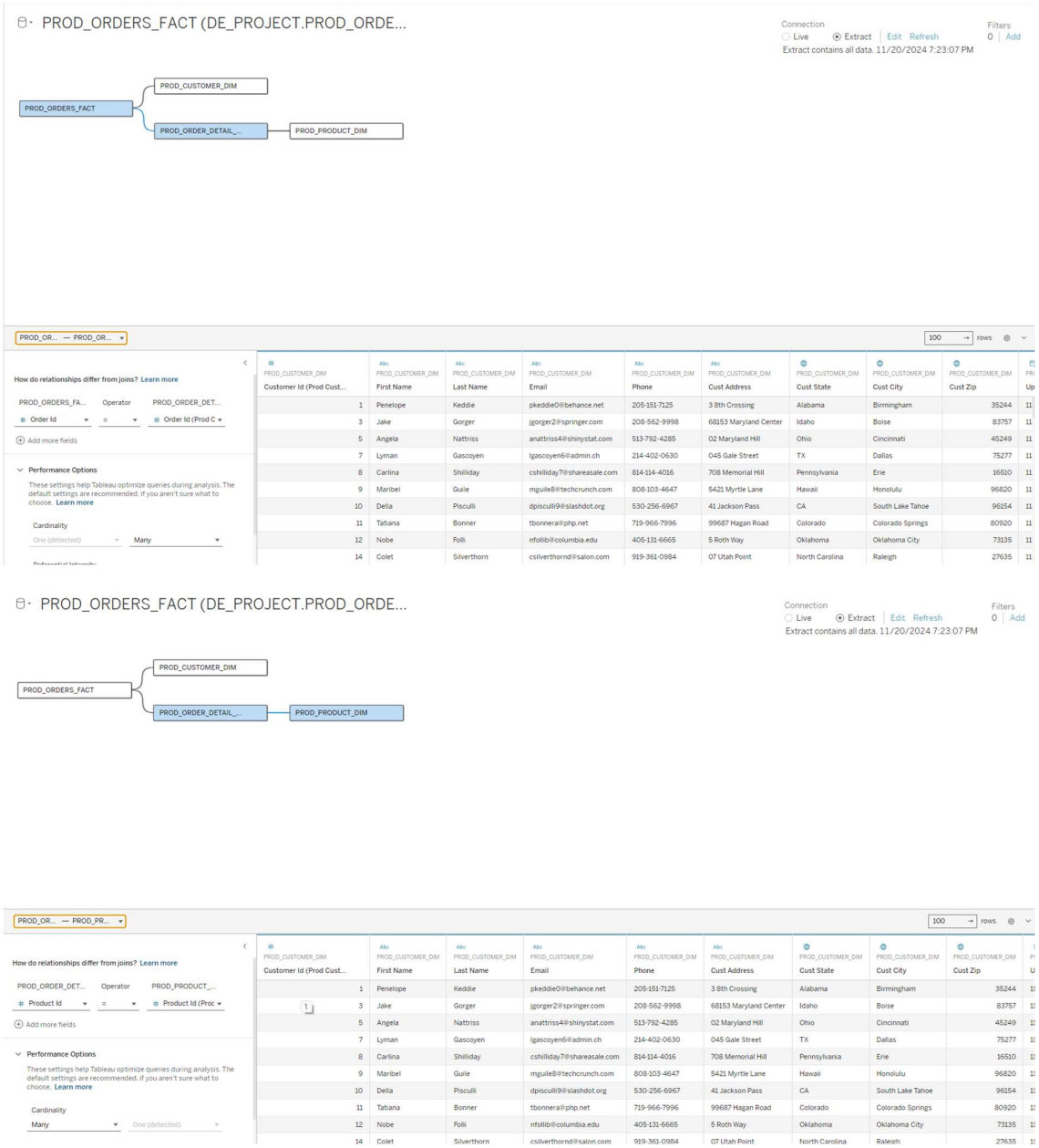
data.



Now I have successfully established a connection between data warehouse and tableau through data extract, not a live connection.

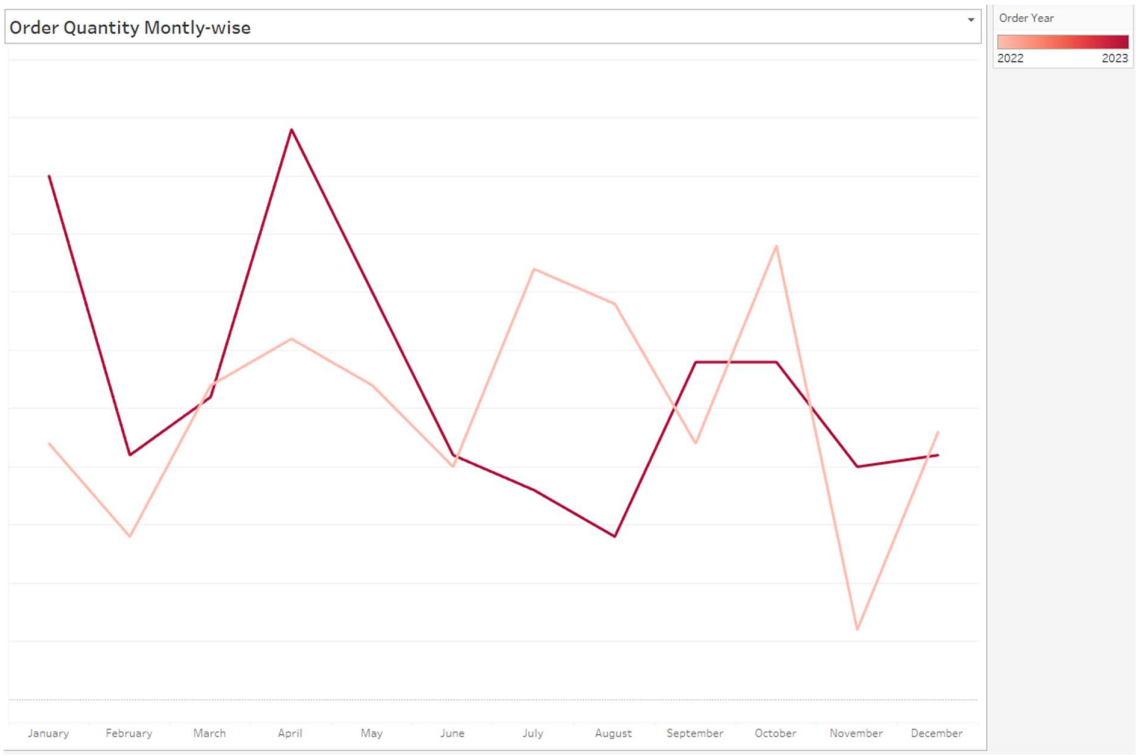




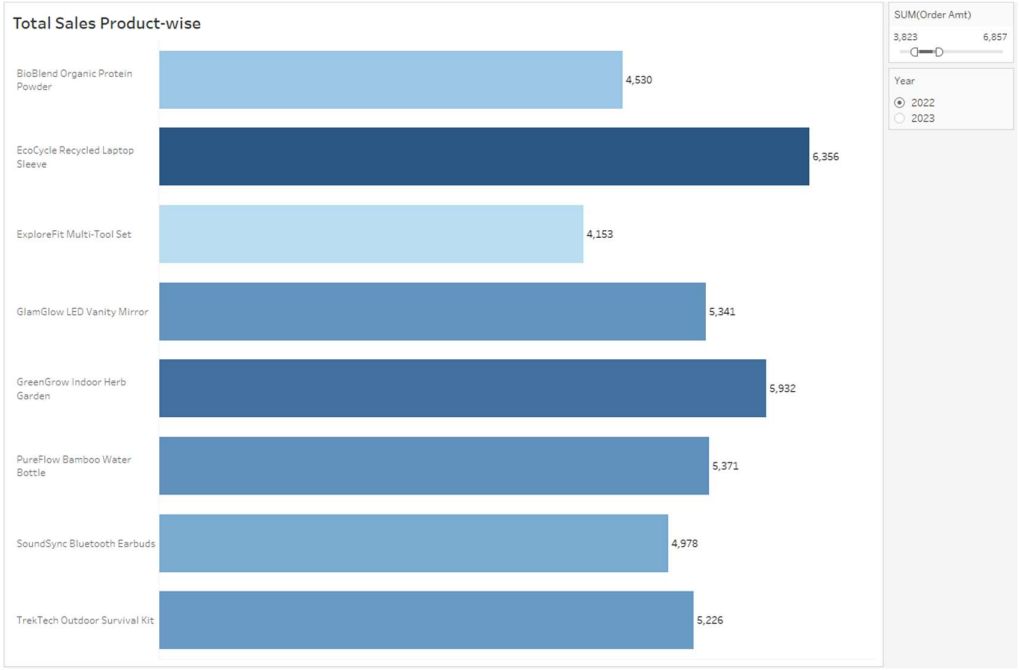


From the above diagram, we could clearly identify that the tables are successfully connected through a foreign key one-to-many relationship.

Visualizations for creating the Dashboard:

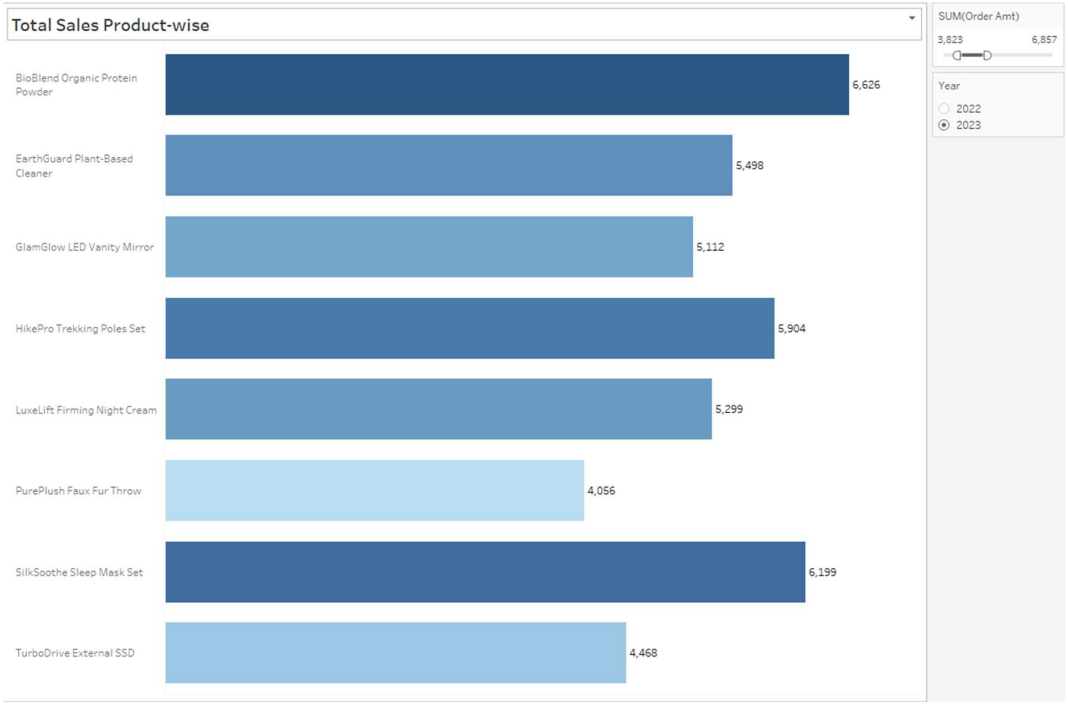


The above line plot shows the unstable trendline of ordered quantities for the years 2022 and 2023, as for the year 2022, orders peaked in the month of October, followed by a downturn in November month. As for the year 2023, the e-commerce company received more orders in the month of April, which could be because of summer discounts.

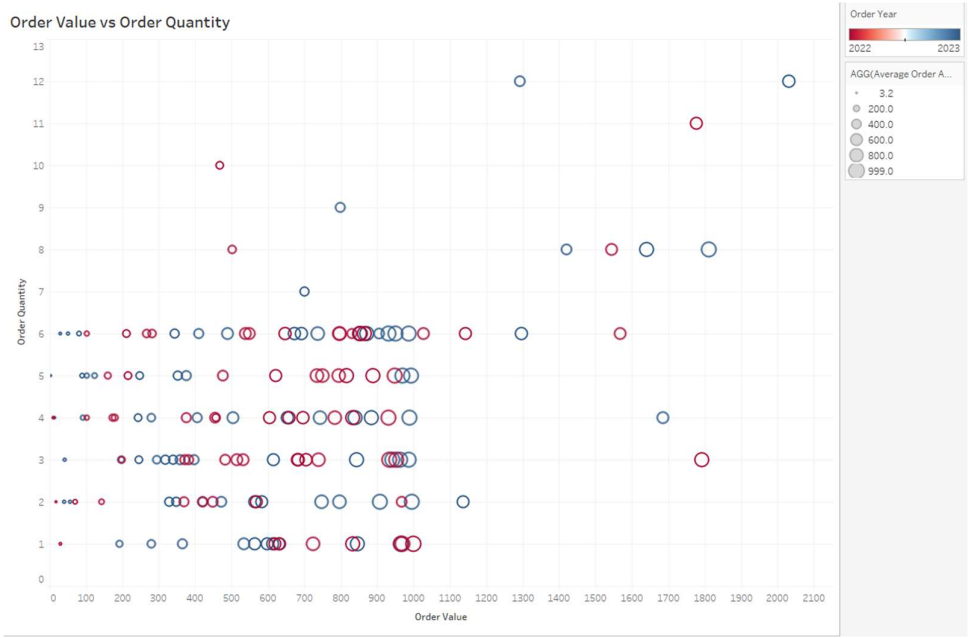


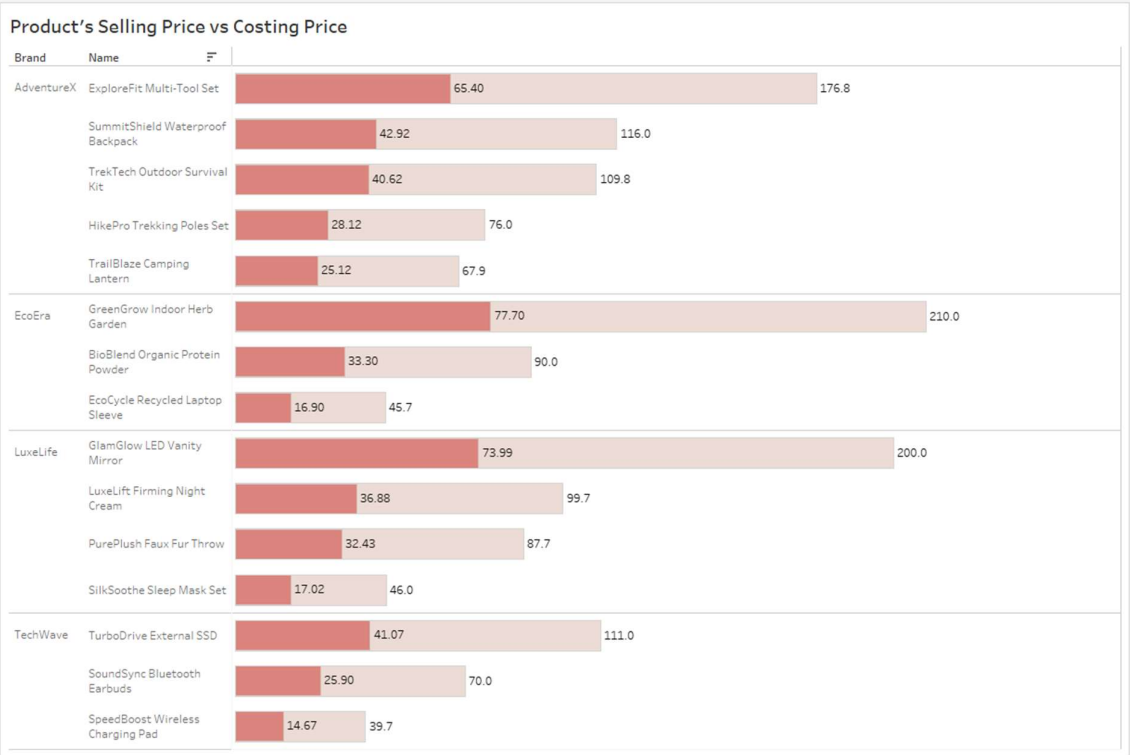
The above bar graphs show us the best-selling product for the year 2022 and the top-selling product in this year is the Eco-cycle Recycled laptop sleeve, contributing a revenue of more than 6000+ USD

As for 2023, below graph shows that Bio-blend organic protein powder had a huge demand with a total sale of around 6,626 USD in amount.

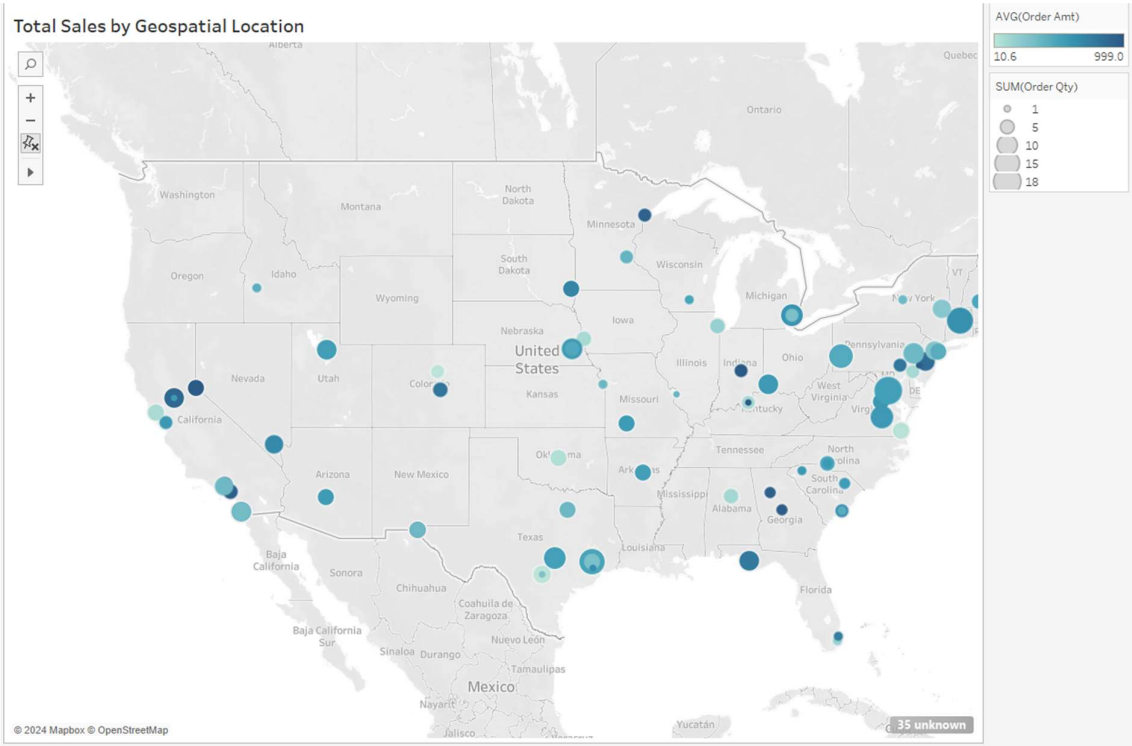


Below scatter plot below shows us the relationship between Order value and Quantity, we notice that most of the order values are cluttered between greater than 0 and lesser than 1100, with the maximum number of orders in terms of quantity and value being placed more in 2023 than 2022.

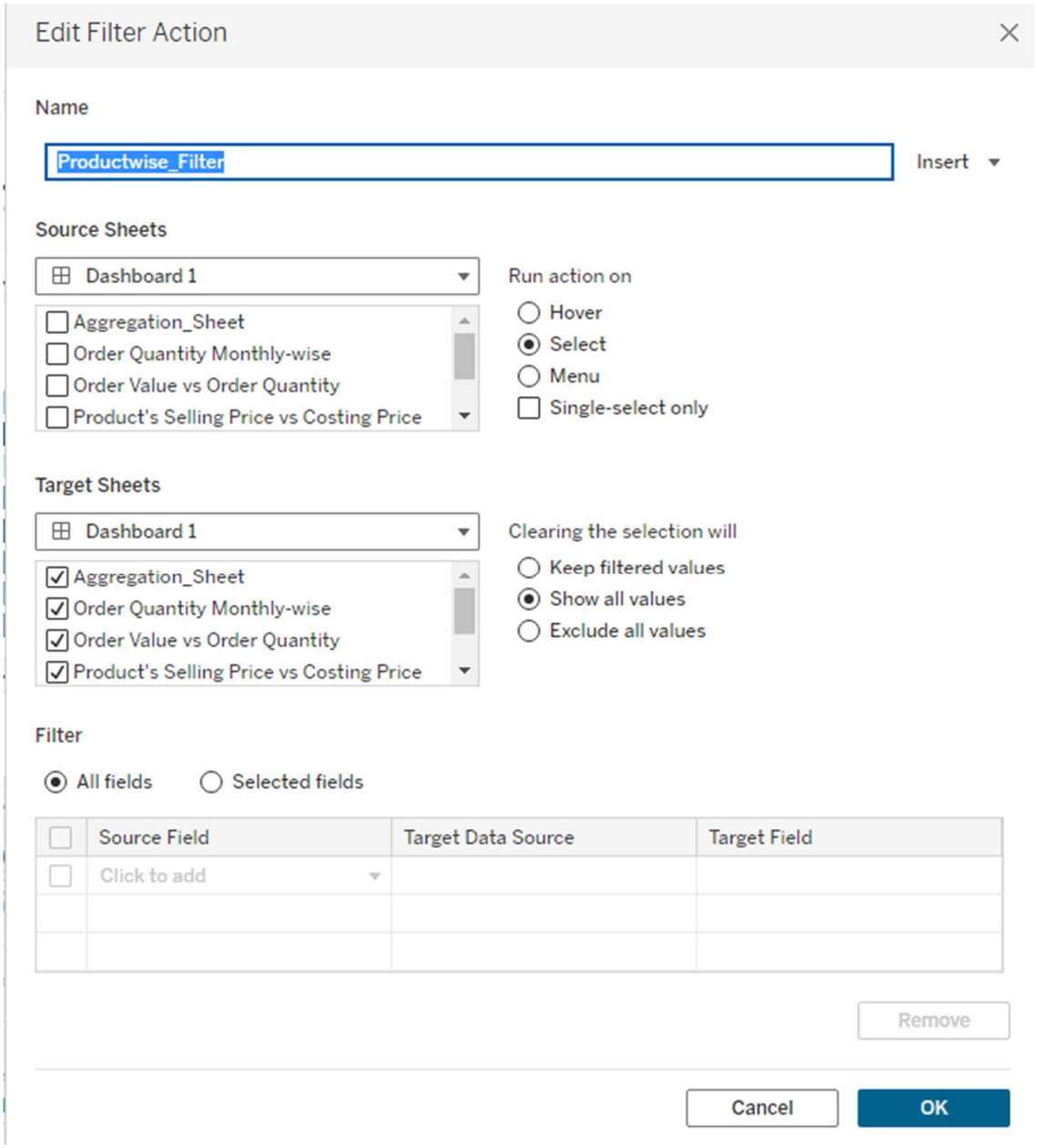


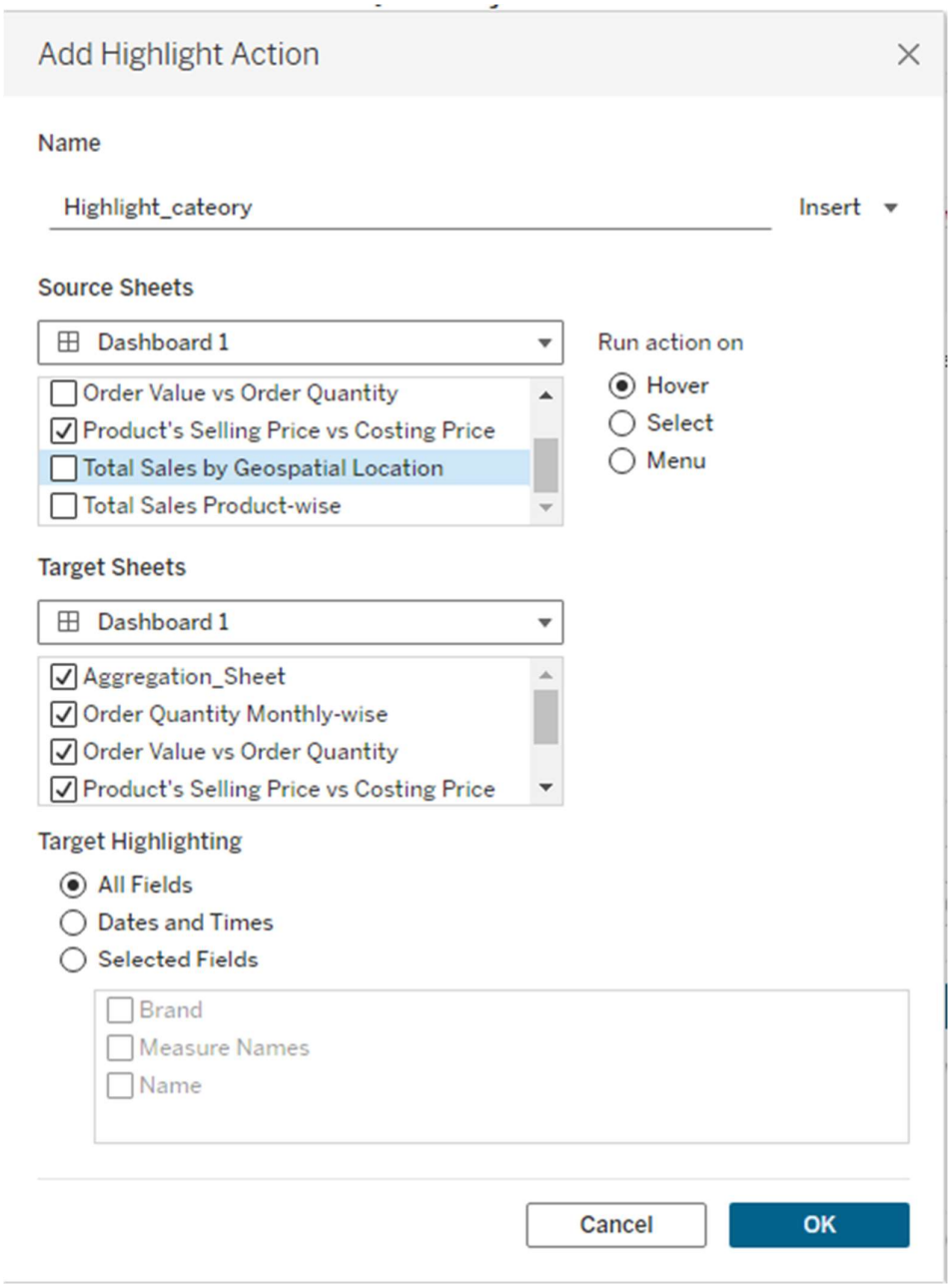


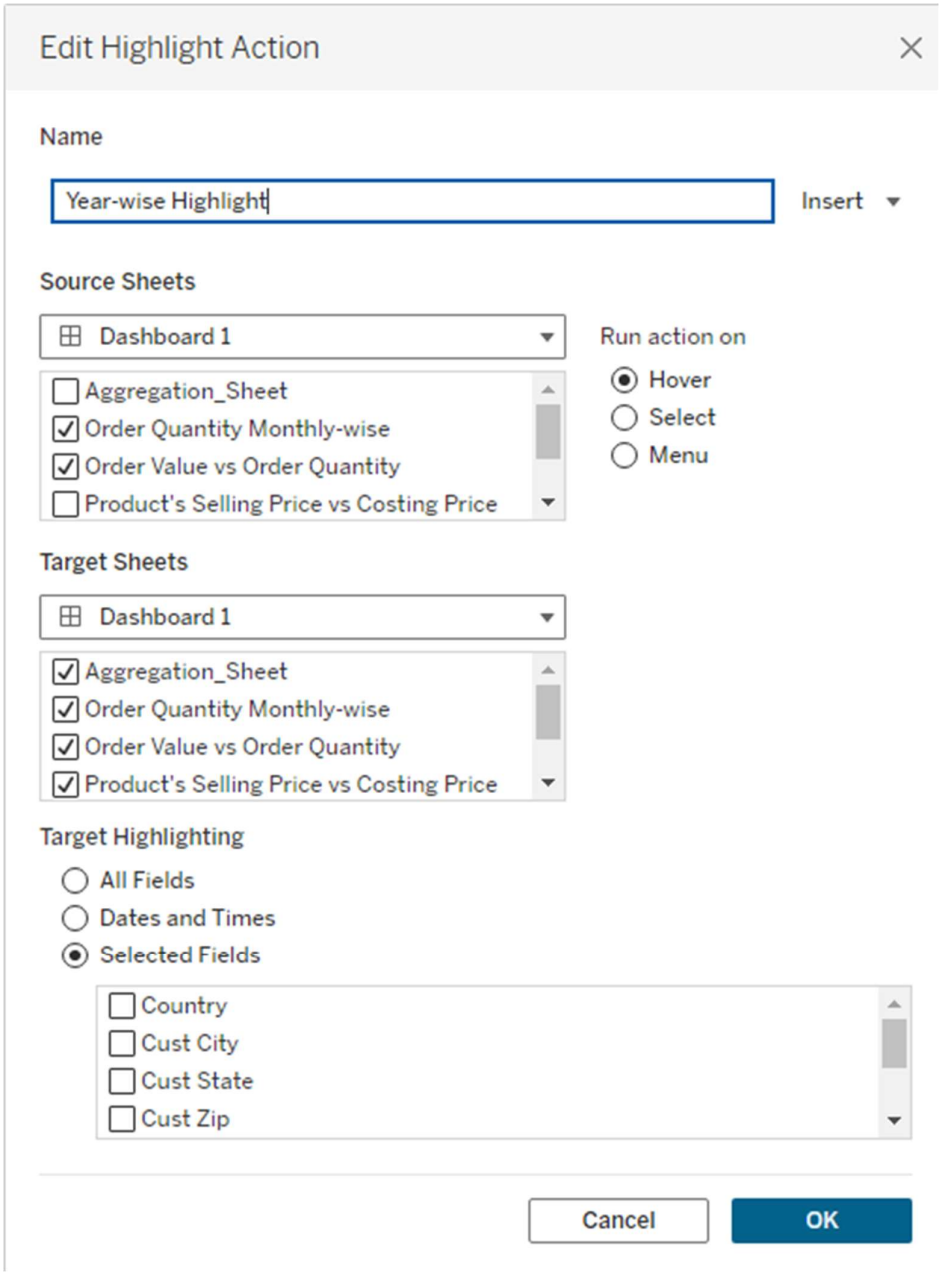
Above stacked bar plot shows us the difference between selling and manufacturing price i.e. costing price for a product, we could notice that all the products are been sold at rate of 2.7x times more than costing price to maintain a profit margin for the company.



As for the e-commerce, most loyal and consistent customers who orders with a huge cart value are mostly around from New-England region specifically, around Massachusetts and Virginia state.


Edit Filter Action
Name
Productwise_Filter
Insert
Source Sheets
Dashboard 1
Aggregation_Sheet
Order Quantity Monthly-wise
Order Value vs Order Quantity
Product's Selling Price vs Costing Price
Run action on
Hover
Select
Menu
Single-select only
Target Sheets
Dashboard 1
Aggregation_Sheet
Order Quantity Monthly-wise
Order Value vs Order Quantity
Product's Selling Price vs Costing Price
Clearing the selection will
Keep filtered values
Show all values
Exclude all values
Filter
All fields
Selected fields
Source Field
Target Data Source
Target Field
Click to add
Remove
Cancel
OK


Add Highlight Action
Name
Highlight_cateory
Insert
Source Sheets
Dashboard 1
Order Value vs Order Quantity
Product's Selling Price vs Costing Price
Total Sales by Geospatial Location
Total Sales Product-wise
Run action on
Hover
Select
Menu
Target Sheets
Dashboard 1
Aggregation_Sheet
Order Quantity Monthly-wise
Order Value vs Order Quantity
Product's Selling Price vs Costing Price
Target Highlighting
All Fields
Dates and Times
Selected Fields
Brand
Measure Names
Name
Cancel
OK

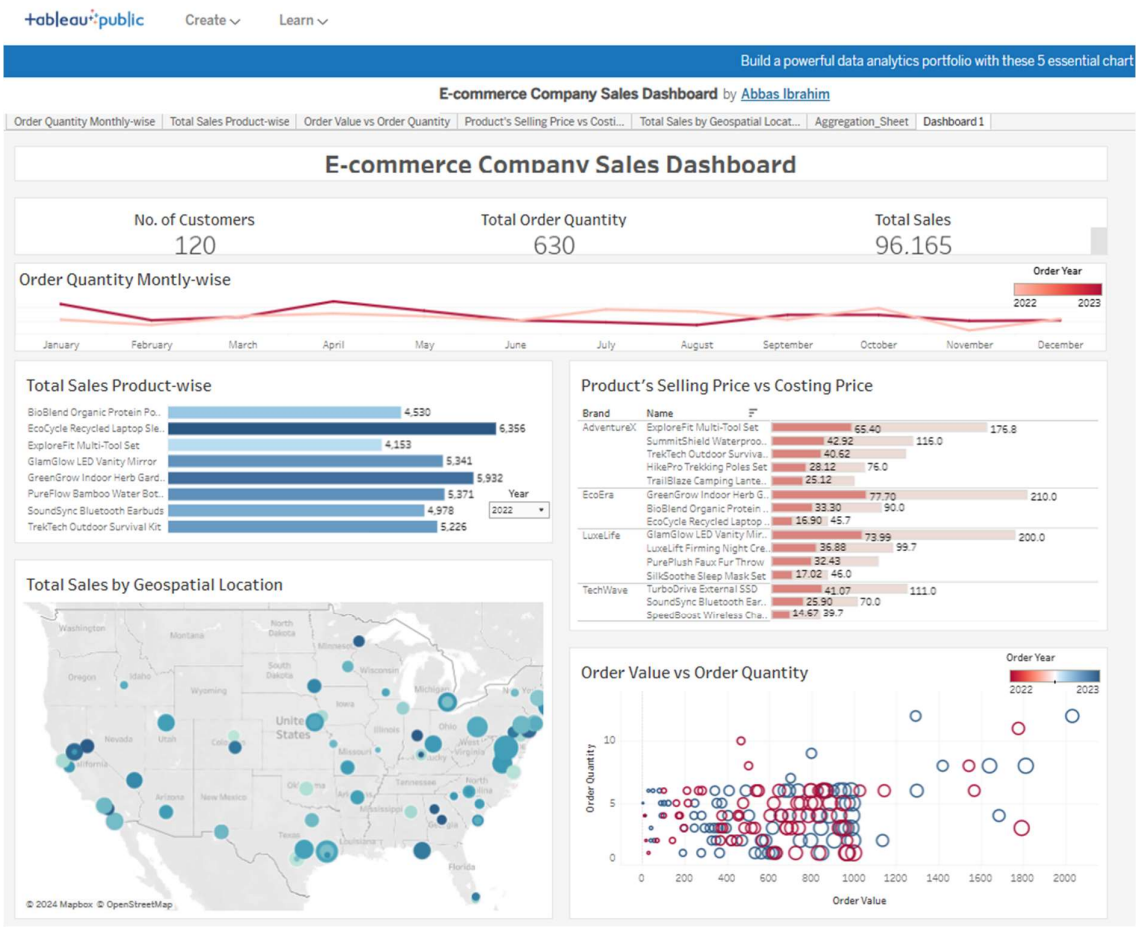


Leveraging tableau, I have created three actions, one is to filter and other two will perform highlight actions.

By consolidating all of the above worksheets into the below dashboard and have uploaded it in the tableau public site.

[E-commerce Company Sales Dashboard | Tableau Public]

https://public.tableau.com/app/profile/abbas.ibrahim3274/viz/E-commerceCompanySalesDashboard/Dashboard1



Projection Reflection:

This project allowed me to visualize, comprehend, and gain hands-on experience in working around the full data engineering lifecycle. Before this project, I never gained real-time experience in implementing and developing an entire data pipeline for the downstream users, but this project started from scratch by creating a synthetic dataset and uploading it to Salesforce (Customer Relationship Manager tool) to mimic the real-time environment, from there utilized an automated Extract, Transform and loading (ETL) tool (Airbyte) to create a connection between Salesforce and Snowflake (Datawarehouse). During this particular process, I faced a specific challenge, where I could not establish a connection between these two using Airbyte because of an error (REST API error), later sorted out this error and successfully established a connection, which allowed me to have a hands-on experience on an ETL tool.

During the second part of our project, I was able to seamlessly create the staging and production schema and tables, then perform data cleaning and processing tasks through SQL and then finally store inside the production table for the downstream users to leverage on the transformed data to extract insights either through Machine learning or a simple visualization.

Finally, I have established a connection between the production schema and Tableau to create Key Performance Indicators (KPI) and several dynamic plots and finally consolidated them into a Dashboard leveraging Tableau Desktop and published it online, which would be ideal and suitable for story-telling it to non-technical stakeholders, thus enabled to convert a raw data into a key driver for taking business decisions.

Reflect upon this project, allowed me to gain hands-on experience and gave me an opportunity to learn new industry-relevant tools and how they can be integrated to implement an entire data engineering lifecycle i.e. from data source generation (Mockeroo & Salesforce), ingestion, transformation, and storage in Snowflake to data visualization in tableau. This project would add great value to my portfolio and would really assist me in exploring relevant roles.

With the skillset that I've learned, I would apply and integrate with other projects to create industry-demanding projects to solve real-world problems that can only be solved through the deep analysis of big data.