


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.



Successfully Loaded Data

TP_Customer_Dim.csv → DE_PROJECT.STAGE_DATA.CUSTOMER_DIM

200 rows were successfully inserted into the table.

Query Data

Done

	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.

197
198
199
200
201
202
203
204
205

SELECT * FROM DE_PROJECT.PROD_DATA.PROD_CUSTOMER_DIM

	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@elashdot.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	nfolli@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.

FileDataServerWindowHelp

Connections

q76539 ca-c...computing.com

Warehouse

COMPUTE_WH

Database

DE_PROJECT

Schema

PROD_DATA

Table

PROD_CUST_OMER_DIM

PROD_ORDER_ETAIL_DIM

PROD_ORDER_DERS_FACT

PROD_PROD_DUCT_DIM

New Custom SQL

New Union

New Table Extension

PROD_ORDERS_FACT (DE_PROJECT.PROD_ORDE...

Connection

Live

Extract

Edit

Refresh

Filters

0

Add

Extract will contain all data.

PROD_ORDERS_FACT

PROD_CUSTOMER_DIM

PROD_ORDER_DETAIL...

PROD_PRODUCT_DIM

PROD_ORDERS_FACT

7 fields 250 rows

100 rows

PROD_ORDERS_FACT	PROD_ORDERS_FACT	PROD_ORDERS_FACT	PROD_ORDERS_FACT	PROD_ORDERS_FACT	PROD_ORDERS_FACT	PROD_ORDERS_FACT
Order Id	Order Amt	Order Qty	Customer Id	Order Date	Order Year	Updated Datetime
1	543.500	3	20	7/14/2022	2022	11/13/2024 9:15:12 PM
2	138.620	4	156	9/25/2022	2022	11/13/2024 9:15:12 PM
3	421.680	1	6	7/22/2023	2023	11/13/2024 9:15:12 PM
4	197.810	3	153	3/20/2022	2022	11/13/2024 9:15:12 PM
5	742.610	3	89	10/5/2023	2023	11/13/2024 9:15:12 PM
6	832.690	1	135	3/10/2023	2022	11/13/2024 9:15:12 PM
7	363.860	5	54	4/19/2023	2022	11/13/2024 9:15:12 PM
8	371.740	3	125	4/19/2023	2022	11/13/2024 9:15:12 PM
9	784.620	1	196	2/1/2023	2023	11/13/2024 9:15:12 PM
10	681.180	1	167	12/4/2022	2022	11/13/2024 9:15:12 PM

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

PROD_ORDERS_FACT (DE_PROJECT.PROD_ORDE...

Connection

Live

Extract

Edit

Refresh

Filters

0

Add

Extract contains all data. 11/20/2024 7:23:07 PM

PROD_ORDERS_FACT

PROD_CUSTOMER_DIM

PROD_ORDER_DETAIL...

PROD_PRODUCT_DIM

PROD_OR... → PROD_CU...

100 rows

How do relationships differ from joins? Learn more

PROD_ORDERS_FACT

Operator

PROD_CUSTOMER...

Customer Id

=

Customer Id (P)

Add more fields

Performance Options

These settings help Tableau optimize queries during analysis. The default settings are recommended. If you aren't sure what to choose. Learn more

Cardinality

Many

One (detected)

Referential Integrity

Some records match

Some records match

PROD_CUSTOMER_DIM	PROD_CUSTOMER_DIM	PROD_CUSTOMER_DIM	PROD_CUSTOMER_DIM	PROD_CUSTOMER_DIM	PROD_CUSTOMER_DIM	PROD_CUSTOMER_DIM	PROD_CUSTOMER_DIM	PROD_CUSTOMER_DIM	PROD_CUSTOMER_DIM	PROD_CUSTOMER_DIM
Customer Id (Prod Cust...	First Name	Last Name	Email	Phone	Cust Address	Cust State	Cust City	Cust Zip		
1	Penelope	Keddie	pkeddie@bance.net	209-151-7125	3 8th Crossing	Alabama	Birmingham	35244	11	
3	Jake	Gorger	jgorger2@springer.com	208-562-9998	68153 Maryland Center	Idaho	Boise	83757	11	
5	Angela	Nattriss	anattriss4@shinystat.com	513-792-4285	02 Maryland Hill	Ohio	Cincinnati	45249	11	
7	Lyman	Gascogen	lgascogen6@admin.ch	214-402-0630	045 Gale Street	TX	Dallas	75277	11	
8	Carina	Shilliday	cshilliday7@shareasale.com	814-114-4016	708 Memorial Hill	Pennsylvania	Erie	16520	11	
9	Maribel	Gaile	mgale8@techorunh.com	808-103-4647	5421 Myrtle Lane	Hawaii	Honolulu	96820	11	
10	Della	Pisculli	dpisculli9@slashdot.org	530-256-6967	41 Jackson Pass	CA	South Lake Tahoe	96154	11	
11	Tatiana	Bonner	tbonnera@jnp.net	719-966-7996	99687 Hagan Road	Colorado	Colorado Springs	80920	11	
12	Nobe	Folli	nfolli@columnia.edu	405-131-6665	5 Roth Way	Oklahoma	Oklahoma City	73135	11	
14	Colet	Silverthorn	csilverthorn@salon.com	919-361-0984	07 Utah Point	North Carolina	Raleigh	27635	11	
18	Dallon	Tewelson	dtewelson@ioa.ic.gov.au	843-103-6405	48 Welch Center	South Carolina	Florence	29505	11	

PROD_ORDERS_FACT (DE_PROJECT.PROD_ORDE...

Connection: Live Extract Edit Refresh Filters: 0 | Add
Extract contains all data. 11/20/2024 7:23:07 PM

PROD_ORDERS_FACT

PROD_CUSTOMER_DIM

PROD_ORDER_DETAIL_...

PROD_PRODUCT_DIM

PROD_OR... → PROD_OR...

How do relationships differ from joins? Learn more

PROD_ORDERS_FACT Operator PROD_ORDER_DETAIL_...
Order Id = Order Id (Prod C

Performance Options

Cardinality: One (detected) Many

PROD_CUSTOMER_DIM Customer Id (Prod Cust...	PROD_CUSTOMER_DIM First Name	PROD_CUSTOMER_DIM Last Name	PROD_CUSTOMER_DIM Email	PROD_CUSTOMER_DIM Phone	PROD_CUSTOMER_DIM Cust Address	PROD_CUSTOMER_DIM Cust State	PROD_CUSTOMER_DIM Cust City	PROD_CUSTOMER_DIM Cust Zip	
1	Penelope	Keddie	pkeddie0@behance.net	205-551-7125	3 8th Crossing	Alabama	Birmingham	35244	11
3	Jake	Gorger	jjgorger2@springer.com	208-562-9998	68153 Maryland Center	Idaho	Boise	83757	11
5	Angela	Nattriss	anattriss4@shinystat.com	513-792-4285	02 Maryland Hill	Ohio	Cincinnati	45249	11
7	Lyman	Gascoyen	lgascoyen6@admin.ch	214-402-0630	045 Gale Street	TX	Dallas	75277	11
8	Carlina	Shilday	cshilday7@shareasale.com	814-114-4016	708 Memorial Hill	Pennsylvania	Erie	16500	11
9	Maribel	Guile	mguille8@techcrunch.com	808-303-4647	5421 Myrtle Lane	Hawaii	Honolulu	96820	11
10	Della	Pisculli	dpisculli9@slashdot.org	530-256-6967	41 Jackson Pass	CA	South Lake Tahoe	96154	11
11	Tatiana	Bonner	tbonnera@php.net	719-966-7996	99687 Hagan Road	Colorado	Colorado Springs	80920	11
12	Nobe	Folli	nfolli@columbia.edu	405-131-6665	5 Roth Way	Oklahoma	Oklahoma City	73135	11
14	Colet	Silverthorn	csilverthornd@salon.com	919-361-0984	07 Utah Point	North Carolina	Raleigh	27635	11

PROD_ORDERS_FACT (DE_PROJECT.PROD_ORDE...

Connection: Live Extract Edit Refresh Filters: 0 | Add
Extract contains all data. 11/20/2024 7:23:07 PM

PROD_ORDERS_FACT

PROD_CUSTOMER_DIM

PROD_ORDER_DETAIL_...

PROD_PRODUCT_DIM

PROD_OR... → PROD_PR...

How do relationships differ from joins? Learn more

PROD_ORDER_DETAIL_... Operator PROD_PRODUCT_...
Product Id = Product Id (Prod C

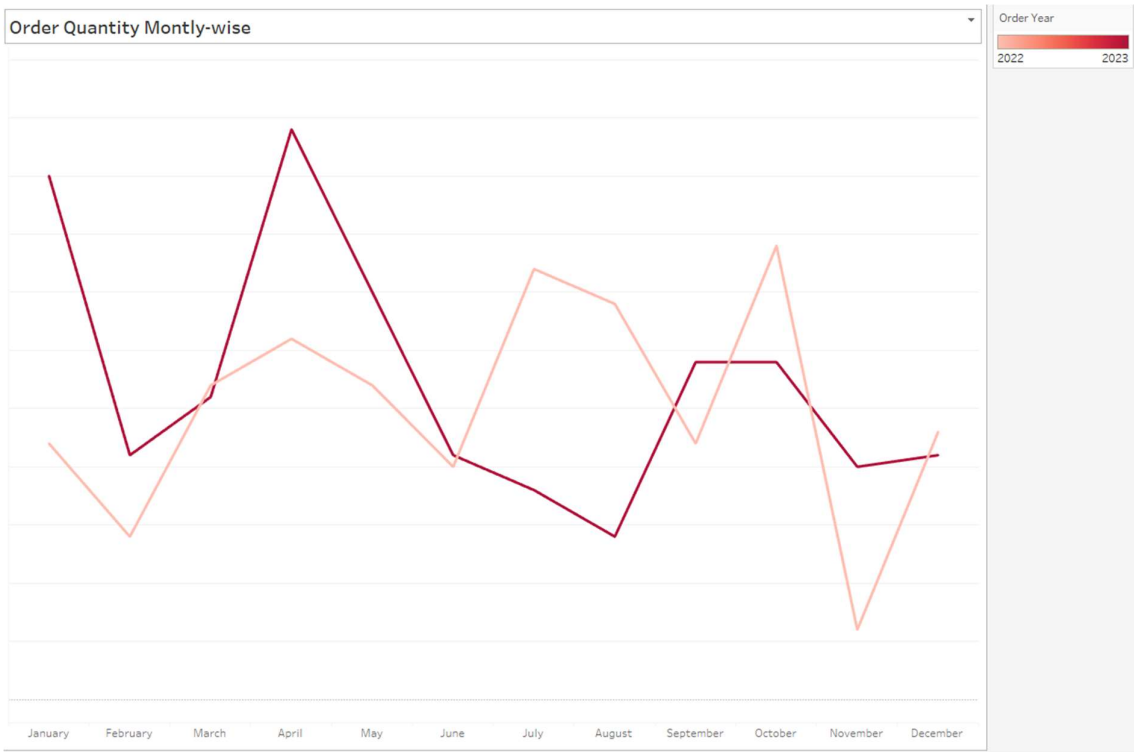
Performance Options

Cardinality: Many One (detected)

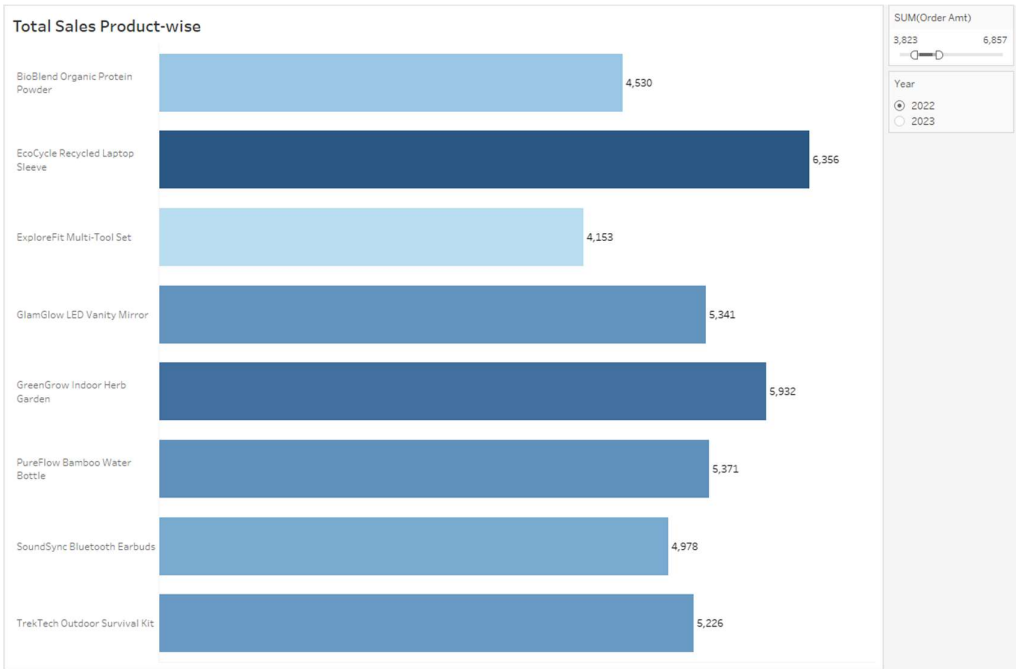
PROD_CUSTOMER_DIM Customer Id (Prod Cust...	PROD_CUSTOMER_DIM First Name	PROD_CUSTOMER_DIM Last Name	PROD_CUSTOMER_DIM Email	PROD_CUSTOMER_DIM Phone	PROD_CUSTOMER_DIM Cust Address	PROD_CUSTOMER_DIM Cust State	PROD_CUSTOMER_DIM Cust City	PROD_CUSTOMER_DIM Cust Zip	
1	Penelope	Keddie	pkeddie0@behance.net	205-551-7125	3 8th Crossing	Alabama	Birmingham	35244	11
3	Jake	Gorger	jjgorger2@springer.com	208-562-9998	68153 Maryland Center	Idaho	Boise	83757	11
5	Angela	Nattriss	anattriss4@shinystat.com	513-792-4285	02 Maryland Hill	Ohio	Cincinnati	45249	11
7	Lyman	Gascoyen	lgascoyen6@admin.ch	214-402-0630	045 Gale Street	TX	Dallas	75277	11
8	Carlina	Shilday	cshilday7@shareasale.com	814-114-4016	708 Memorial Hill	Pennsylvania	Erie	16500	11
9	Maribel	Guile	mguille8@techcrunch.com	808-303-4647	5421 Myrtle Lane	Hawaii	Honolulu	96820	11
10	Della	Pisculli	dpisculli9@slashdot.org	530-256-6967	41 Jackson Pass	CA	South Lake Tahoe	96154	11
11	Tatiana	Bonner	tbonnera@php.net	719-966-7996	99687 Hagan Road	Colorado	Colorado Springs	80920	11
12	Nobe	Folli	nfolli@columbia.edu	405-131-6665	5 Roth Way	Oklahoma	Oklahoma City	73135	11
14	Colet	Silverthorn	csilverthornd@salon.com	919-361-0984	07 Utah Point	North Carolina	Raleigh	27635	11

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:

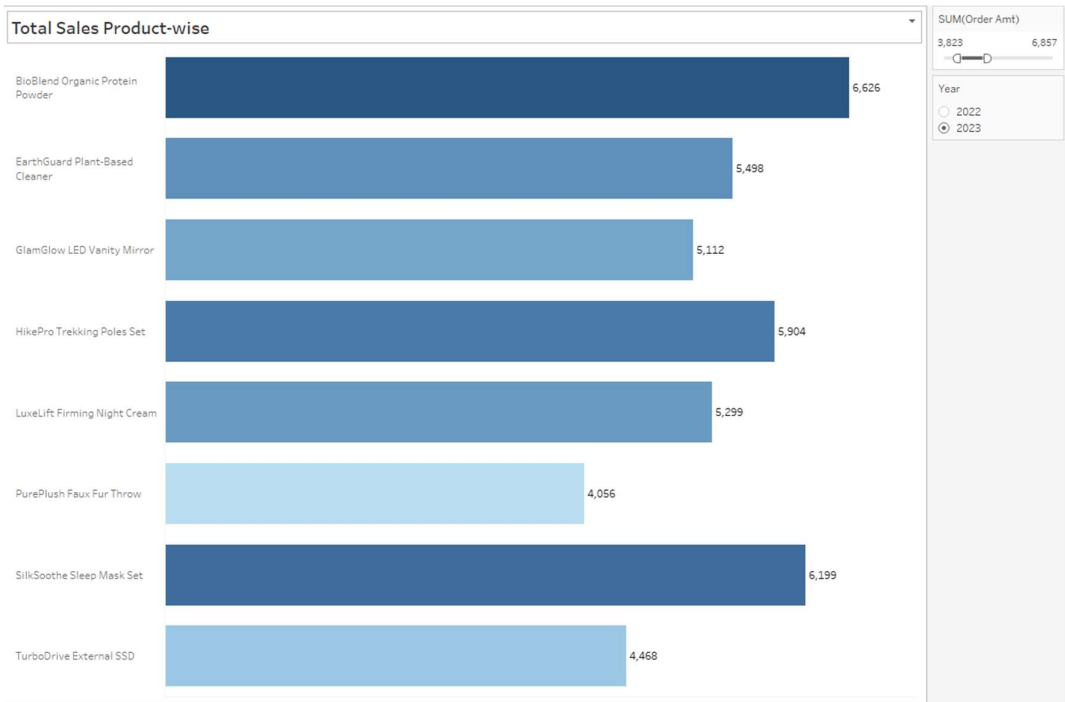


The above line plot shows the unstable trendline of ordered quantities for the years 2022 and 2023, as for the year 2022, orders peaked in the month of October, followed by a downturn in November month. As for the year 2023, the e-commerce company received more orders in the month of April, which could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

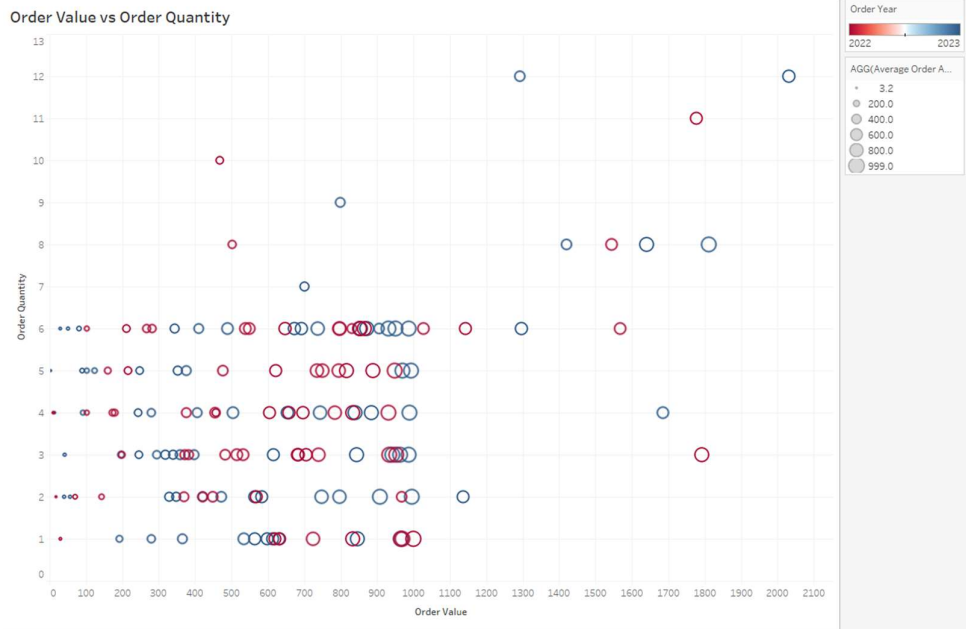


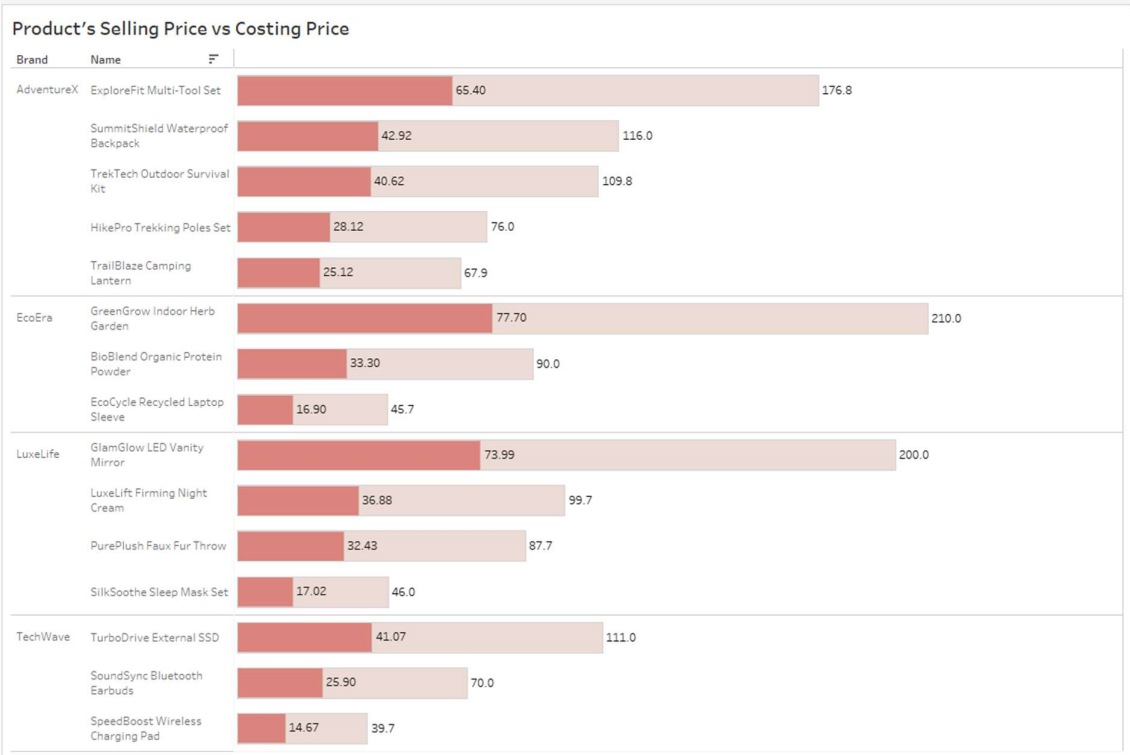
The above bar graphs show us the best-selling product for the year 2022 and the top-selling product in this year is the Eco-cycle Recycled laptop sleeve, contributing a revenue of more than 6000+ USD

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.

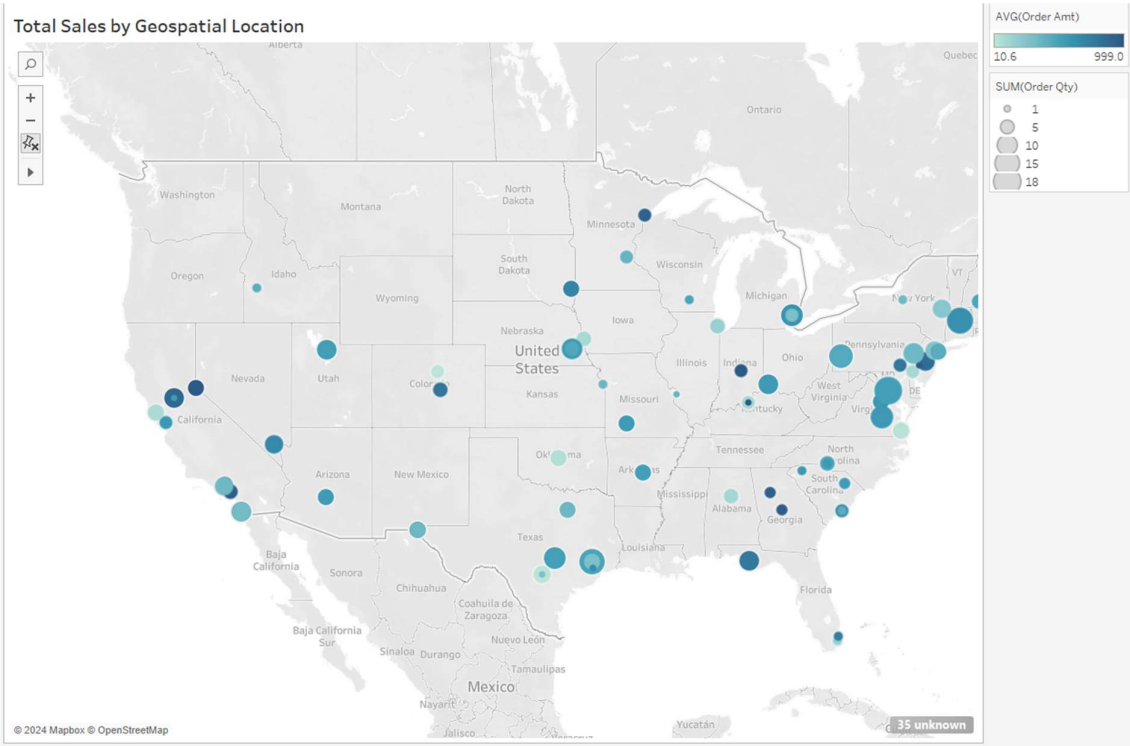


Below scatter plot below shows us the relationship between Order value and Quantity, we notice that most of the order values are cluttered between greater than 0 and lesser than 1100, with the maximum number of orders in terms of quantity and value being placed more in 2023 than 2022.





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

<input type="checkbox"/>	Source Field	Target Data Source	Target Field
<input type="checkbox"/>	Click to add ▾		

Remove

Cancel

OK

Add Highlight Action



Name

Highlight_category

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

Edit Highlight Action

×

Name

Year-wise Highlight

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

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

☐ Country

☐ Cust City

☐ Cust State

☐ Cust Zip

Cancel

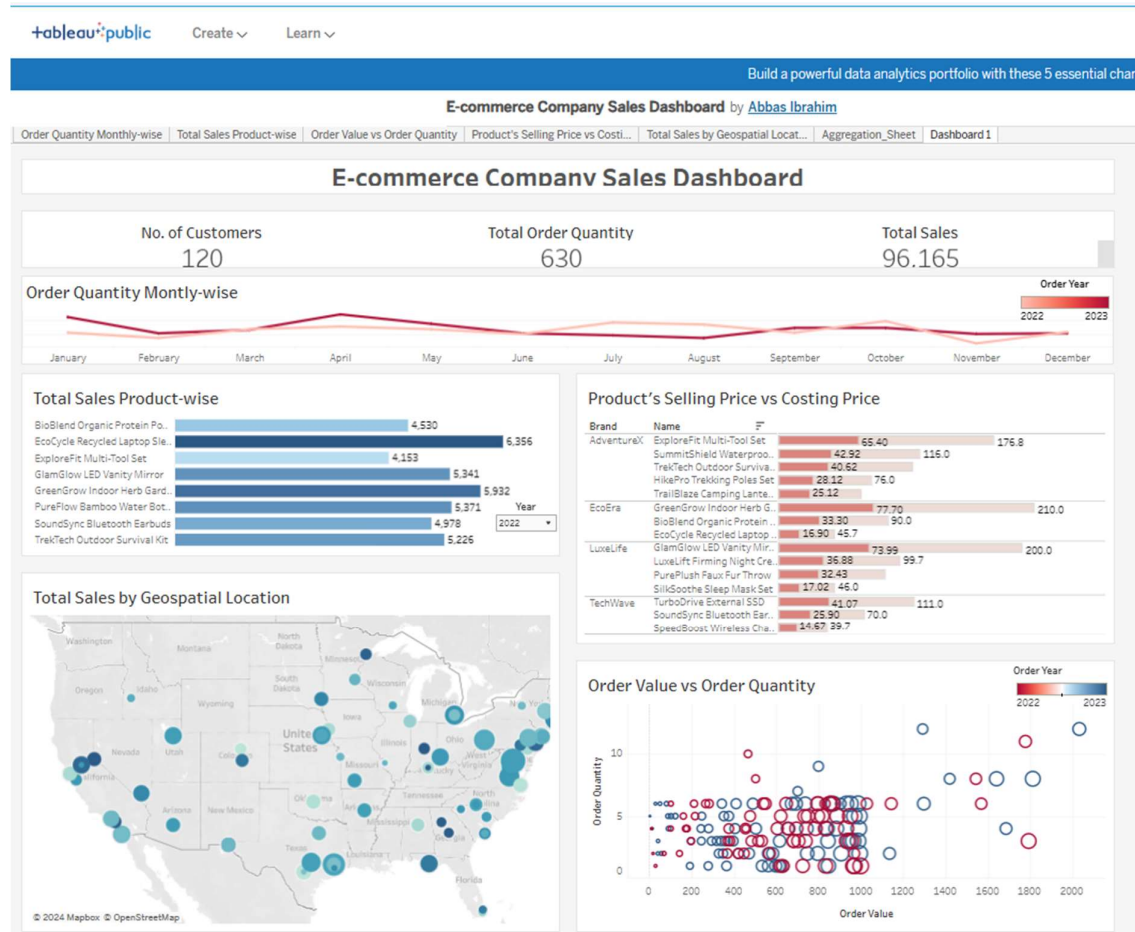
OK

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)

<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 (Mockaroo & 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.