NAAN MUDHALVAN – PROFESSIONAL READINESS FOR INNOVATION, EMPLOYMENT AND ENTERPRENEURSHIP

DATA ANALYTICS ASSIGNMENT – 1

ASSIGNMENT DATE	28-04-23
STUDENT NAME	Aravind.S
STUDENT ROLL NO	820420205011

The growth of supermarkets in most populated cities is increasing and market competitions are also, high. The dataset is one of the historical sales of supermarket company which has recorded in 3 different branches for 3 months data. Predictive data analytics methods are easy to apply to this dataset.

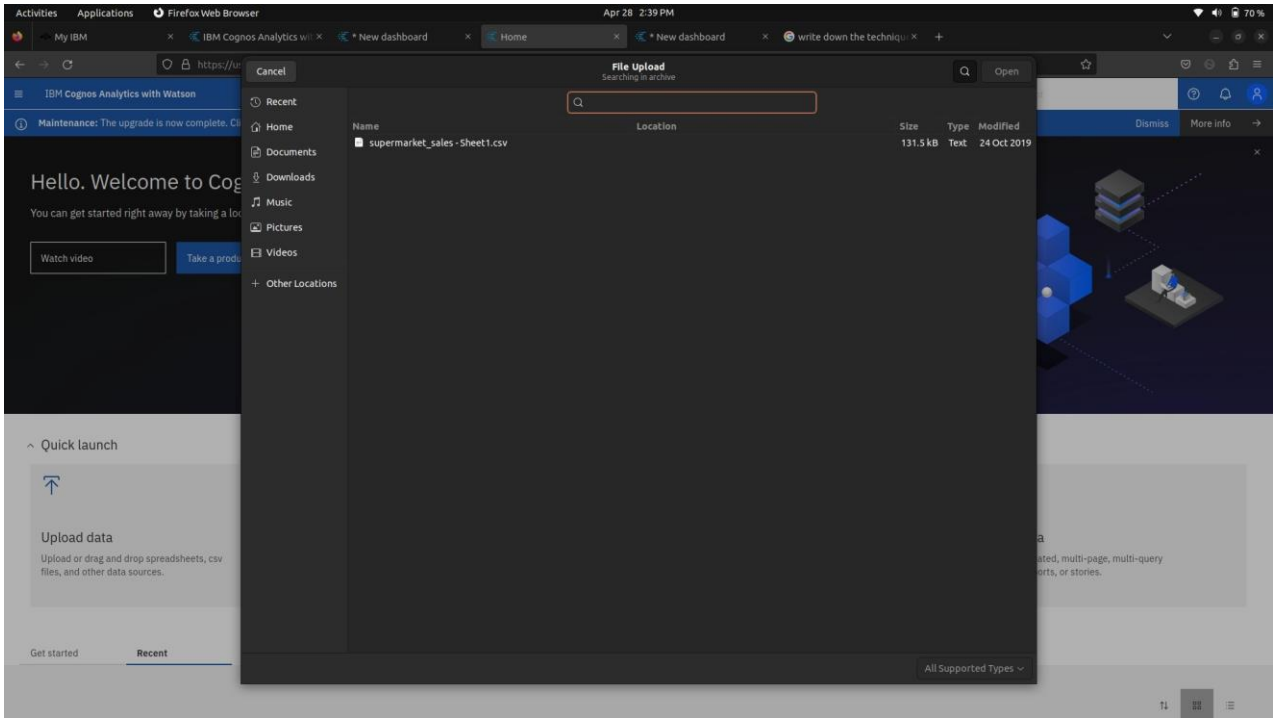
Attribute information:

Invoice id: Computer-generated sales slip invoice identification number
Branch: Branch of supercenter (3 branches are available identified by A, B and C).
City: Location of supercenters
Customer type: Type of customers, recorded by Members for customers using member cards and Normal for those without member cards.
Gender: Gender type of customer
Product line: General item categorization groups - Electronic accessories, Fashion accessories, Food and beverages, Health and beauty, Home and lifestyle, Sports and travel
Unit price: The price of each product in $
Quantity: Number of products purchased by the customer
Tax: 5% tax fee for customers buying
Total: Total price including tax
Date: Date of purchase (Record available from January 2019 to March 2019)
Time: Purchase time (10 am to 9 pm)
Payment: Payment used by the customer for the purchase (3 methods are available – Cash, Credit card and E-wallet)
COGS: Cost of goods sold
Gross margin percentage: Gross margin percentage
Gross income: Gross income
Rating: Customer stratification rating on their overall shopping experience (On a scale of 1 to 10)
Dataset Link: Dataset

There are three steps to be followed in this assignment

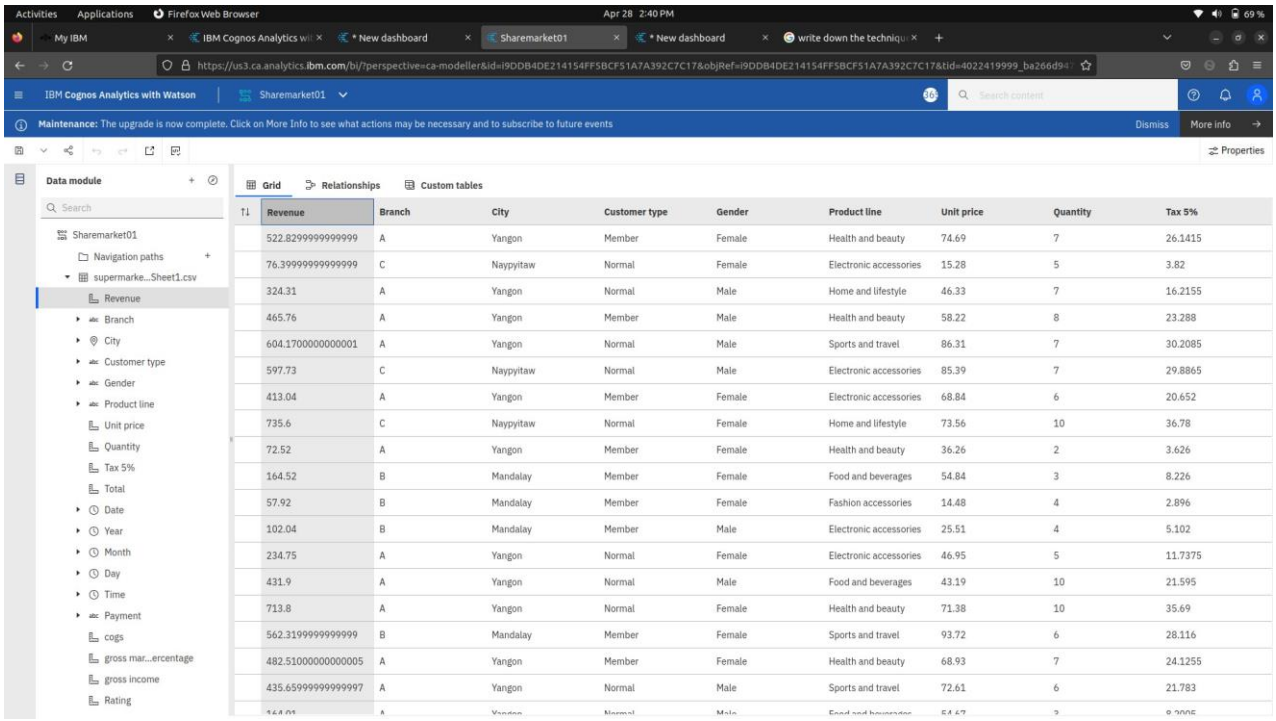
- Upload the dataset into IBM Cognizant.
- Deleting unnecessary data and creating the data module.
- Explore and visualize the dataset.

Step 1: The dataset has been uploaded into IBM cognizant successfully after downloading the given dataset.



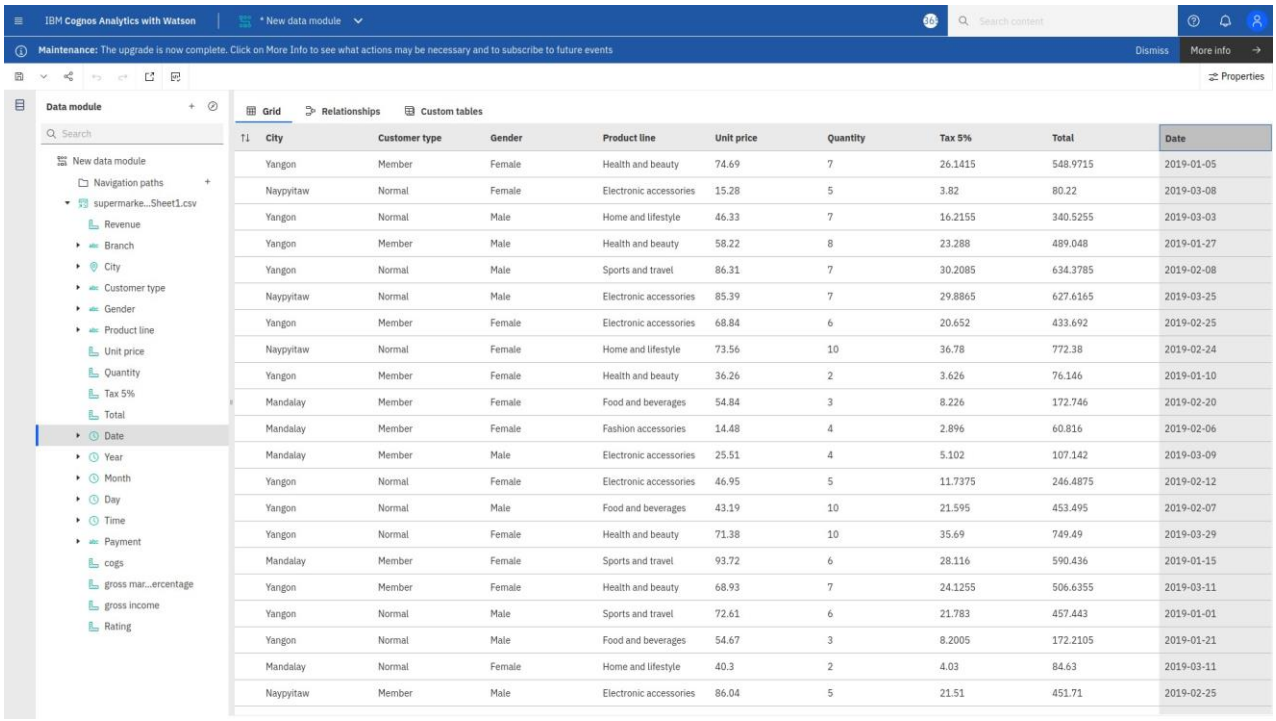
Step 2: Cleaning the dataset using “Prepare Data”. I splitted some of the columns, so it can be used for better visualization of dataset. I also added some columns using existing columns and manipulated it using calculations.

Here we can see the newly added column called Revenue.



Revenue	Branch	City	Customer type	Gender	Product line	Unit price	Quantity	Tax 5%
522.8299999999999	A	Yangon	Member	Female	Health and beauty	74.69	7	26.1415
76.39999999999999	C	Naypyitaw	Normal	Female	Electronic accessories	15.28	5	3.82
324.31	A	Yangon	Normal	Male	Home and lifestyle	46.33	7	16.2155
465.76	A	Yangon	Member	Male	Health and beauty	58.22	8	23.288
604.1700000000001	A	Yangon	Normal	Male	Sports and travel	86.31	7	30.2085
597.73	C	Naypyitaw	Normal	Male	Electronic accessories	85.39	7	29.8865
413.04	A	Yangon	Member	Female	Electronic accessories	68.84	6	20.652
735.6	C	Naypyitaw	Normal	Female	Home and lifestyle	73.56	10	36.78
72.52	A	Yangon	Member	Female	Health and beauty	36.26	2	3.626
164.52	B	Mandalay	Member	Female	Food and beverages	54.84	3	8.226
57.92	B	Mandalay	Member	Female	Fashion accessories	14.48	4	2.896
102.04	B	Mandalay	Member	Male	Electronic accessories	25.51	4	5.102
234.75	A	Yangon	Normal	Female	Electronic accessories	46.95	5	11.7375
431.9	A	Yangon	Normal	Male	Food and beverages	43.19	10	21.595
713.8	A	Yangon	Normal	Female	Health and beauty	71.38	10	35.69
562.3199999999999	B	Mandalay	Member	Female	Sports and travel	93.72	6	28.116
482.51000000000005	A	Yangon	Member	Female	Health and beauty	68.93	7	24.1255
435.65999999999997	A	Yangon	Normal	Male	Sports and travel	72.61	6	21.783

Here the date column is splitted into year, month, day and day of the week.



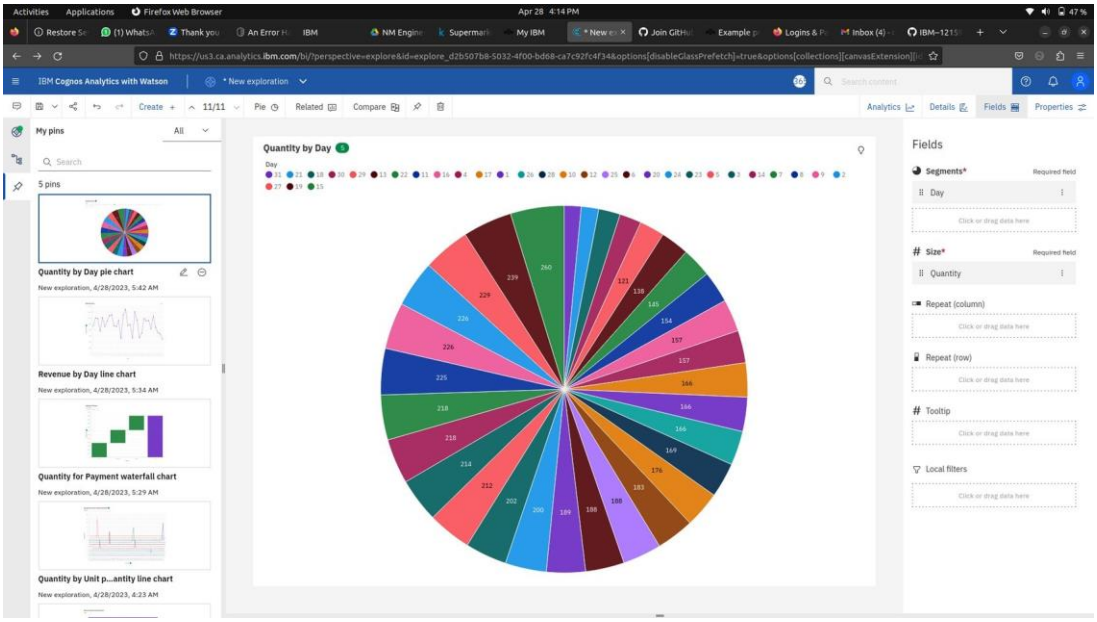
City	Customer type	Gender	Product line	Unit price	Quantity	Tax 5%	Total	Date
Yangon	Member	Female	Health and beauty	74.69	7	26.1415	548.9715	2019-01-05
Naypyitaw	Normal	Female	Electronic accessories	15.28	5	3.82	80.22	2019-03-08
Yangon	Normal	Male	Home and lifestyle	46.33	7	16.2155	340.5255	2019-03-03
Yangon	Member	Male	Health and beauty	58.22	8	23.288	489.048	2019-01-27
Yangon	Normal	Male	Sports and travel	86.31	7	30.2085	634.3785	2019-02-08
Naypyitaw	Normal	Male	Electronic accessories	85.39	7	29.8865	627.6165	2019-03-25
Yangon	Member	Female	Electronic accessories	68.84	6	20.652	433.692	2019-02-25
Naypyitaw	Normal	Female	Home and lifestyle	73.56	10	36.78	772.38	2019-02-24
Yangon	Member	Female	Health and beauty	36.26	2	3.626	76.146	2019-01-10
Mandalay	Member	Female	Food and beverages	54.84	3	8.226	172.746	2019-02-20
Mandalay	Member	Female	Fashion accessories	14.48	4	2.896	60.816	2019-02-06
Mandalay	Member	Male	Electronic accessories	25.51	4	5.102	107.142	2019-03-09
Yangon	Normal	Female	Electronic accessories	46.95	5	11.7375	246.4875	2019-02-12
Yangon	Normal	Male	Food and beverages	43.19	10	21.595	453.495	2019-02-07
Yangon	Normal	Female	Health and beauty	71.38	10	35.69	749.49	2019-03-29
Mandalay	Member	Female	Sports and travel	93.72	6	28.116	590.436	2019-01-15
Yangon	Member	Female	Health and beauty	68.93	7	24.1255	506.6355	2019-03-11
Yangon	Normal	Male	Sports and travel	72.61	6	21.783	457.443	2019-01-01
Yangon	Normal	Male	Food and beverages	54.67	3	8.2005	172.2105	2019-01-21
Mandalay	Normal	Female	Home and lifestyle	40.3	2	4.03	84.63	2019-03-11
Naypyitaw	Member	Male	Electronic accessories	86.04	5	21.51	451.71	2019-02-25

Like this we cleaned the dataset and removed unnecessary information in the dataset.

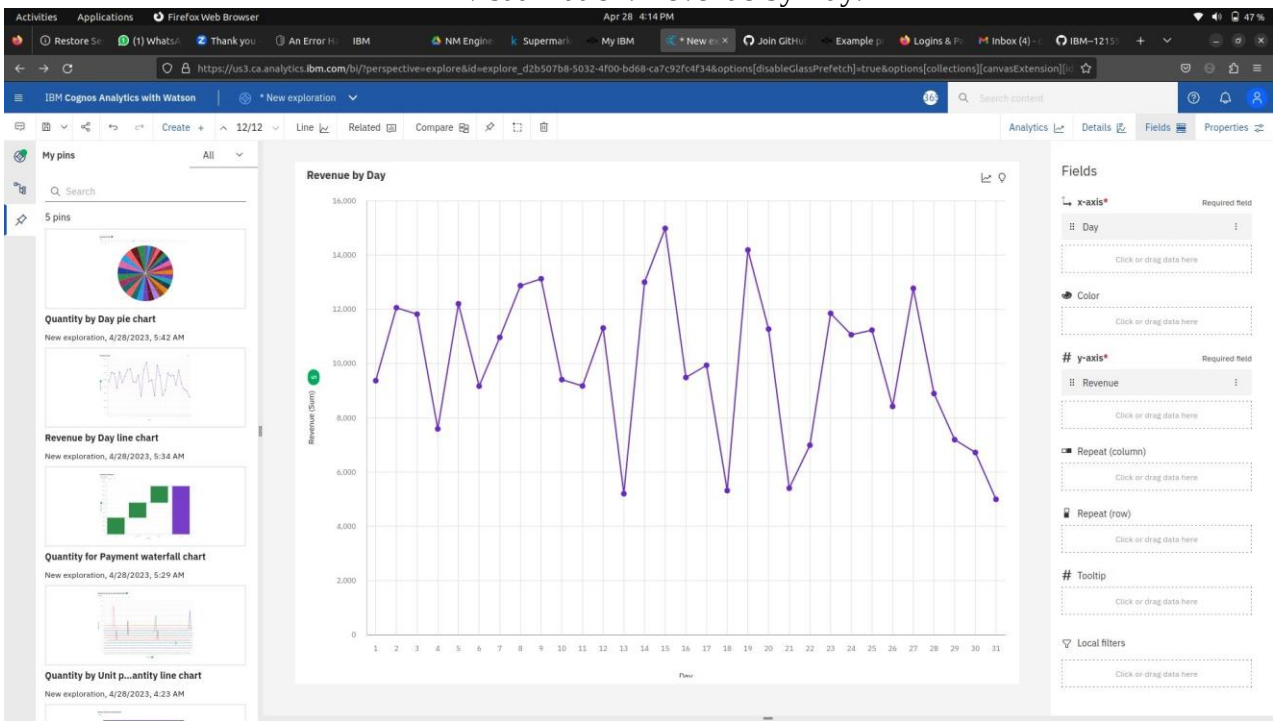
Step **3:** Here we are exploring and visualizing the dataset.

I have done 5 visualizations using Explore data option in IBM Cognizant.

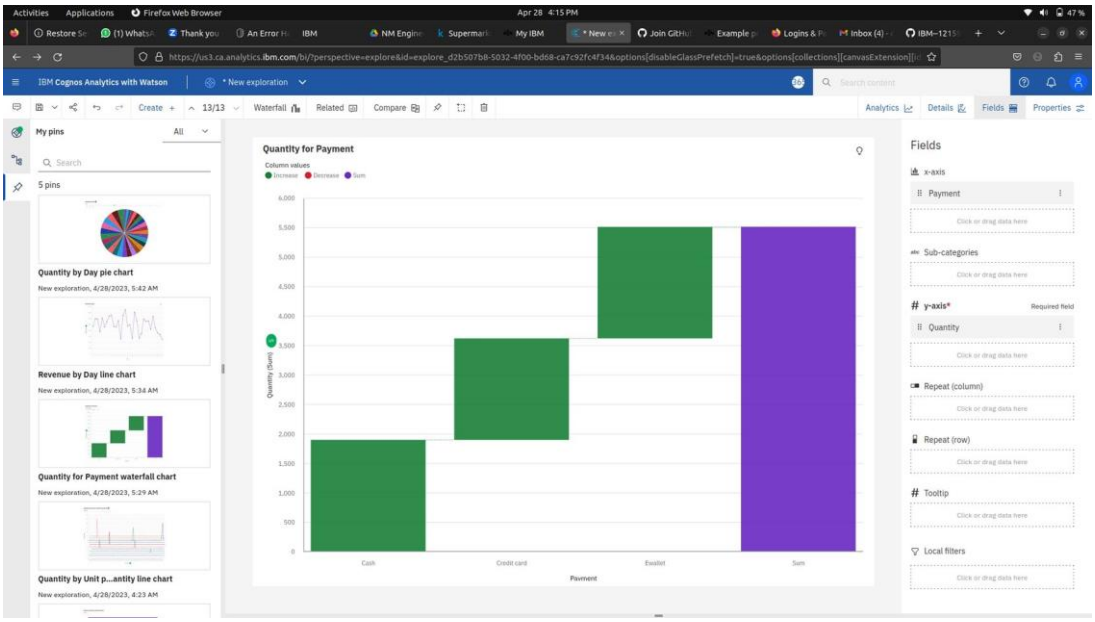
1st visualization: Quantity by Day



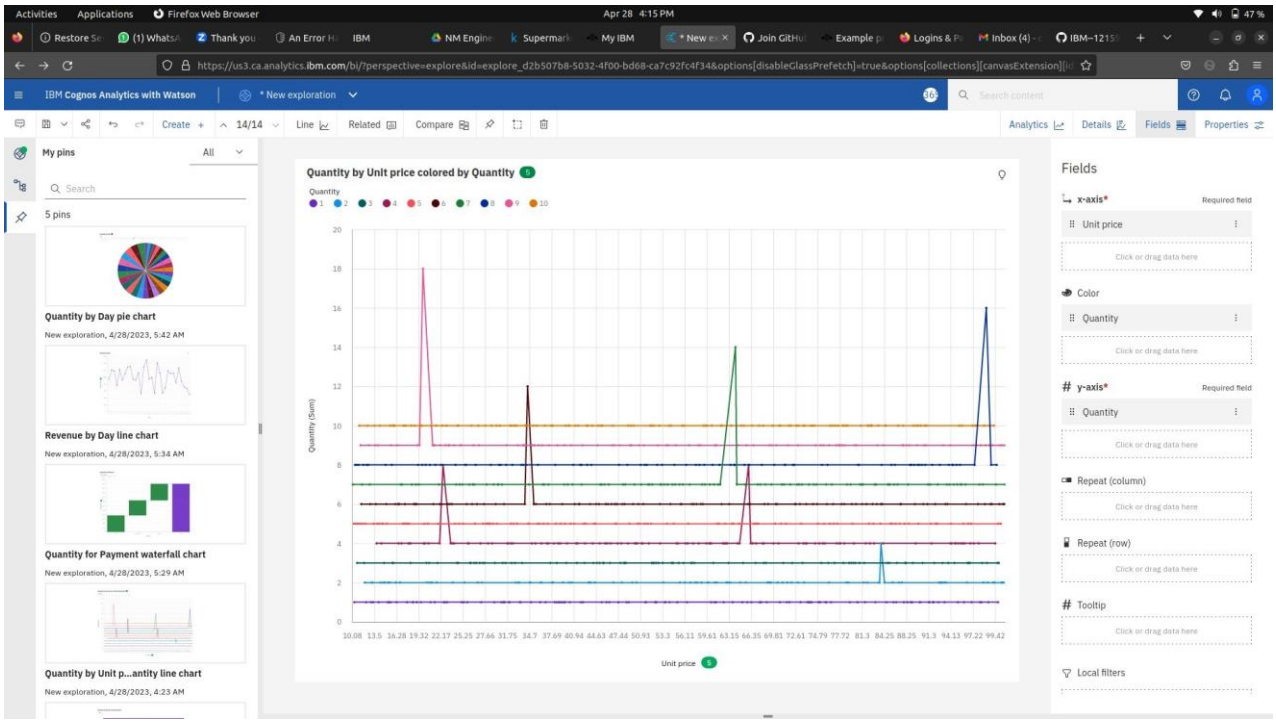
2nd visualization: Revenue by Day.



3rd visualization: Quantity for payment.



4th visualization: Quantity by Unit Price colored By Quantity.



5th visualization: Revenue by Month colored by Month.