**Symbiosis Skills and Professional University**

**Kiwale, Pune**

**PROJECT REPORT**

**On**

**“Food Products Advertising Analysis”**



****

**Submitted by**

**Mr. Vijay Bhagawat Kate**

**(Registration Number: 2201207121)**

**DA-Batch-VI**

**Under The Guidance of Trainer's Name:**  Miss. Vaishnavi Satav.

**STUDENT DECLARATION AND ATTESTATION BY TRAINER**

This is to declare that this report has been written by me. No part of the report is plagiarized from other sources. All information included from other sources have been duly acknowledged. I aver that if any part of the report is found to be plagiarized, I shall take full responsibility for it.

Signature of student

**Name of student: Mr. Vijay Bhagawat Kate** Registration Number:2201207121

Signature of trainer

**Name of trainer:** Miss. Vaishnavi Satav**.**

**CERTIFICATE**

This is to certify that the report entitled,“ **Food Products Advertising Analysis** ”submitted by **” Vijay Bhagawat Kate** ” to Symbiosis Skills and Professional University, Pune, Maharashtra, India, is a record of bonafide Project work carried out by him under my supervision and guidance and is worthy of consideration for the completion of certificate course in ‘Data Associate”.

Signature of Trainer Name of Trainer

Miss. Vaishnavi Satav.

Date: / /

Supervisor Supervisor

Date:

**ACKNOWLEDGEMENTS**

This project report has been prepared under the guidance of our Trainers Mr. Kushal Sharma and Miss. Vaishnavi Satav. A project is a bridge between theoretical and practical learning and with this thinking i worked on the project and made it successful due to timely support and efforts of all who helped me.

I would like to express my profound gratitude towards them for their guidance and timely support throughout the completion of this project.

Also, i would like to thank my team members who worked with me on this project for their dedication and hard work. And friends for their support, suggestions, and guidance for this project.

**INDEX**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Sl. No** | **Index** |  |  |  | **Page**  **No.** |
| 1 | **Acknowledgement** | |  |  |  |
| 2 | **Plan of Project** | |  |  |  |
|  | 2.1 | Purpose of Project | |  | 6 |
|  |  |  | |  |  |
| 3 | **Problem Statements** | | |  | 7 |
| 4 | **Objective of the Project** | | |  |  |
|  | 4.1 | Objectives | |  | 8 |
|  |  |  | | |  |
| 5 | **Data Analysis** | | |  |  |
|  | 5.1 | Tools used | |  | 9 |
|  | 5.2 | Data Cleaning | |  | 10 |
|  | 5.3 | Exploratory Data Analysis | |  | 13 |
|  | 5.4 | Machine Learning | |  | 23 |
|  | 5.5 | Data Visualization | |  | 28 |
|  | 5.6 | Partitioning of data | |  | 29 |
| 6 | **Results** | | |  | 30 |
| 7 | **Conclusion.** | | |  | 31 |

**Purpose of Project:**

Food Mart (CFM) is a chain of convenience stores in the United States as well as other countries. The private company's headquarters are located in main cities of the countries, approximately 325 stores are located in the US Convenient Food Mart operates on the franchise system.  
Food Mart was the nation's third-largest chain of convenience stores as of 1988.

With the help of data that we have, its about the customers that running food marts and their related information USA, Canada, Mexico countries. Our purpose is to do exploratory data analysis and to find meaningful information from data that helps franchise to take useful decisions and to devise Machine Learning Models that helps them predict the cost of media campaigns in the food marts for increasing their sale on the basis of the features provided.

**Problem Statements:**

1. To get details of sale, cost and profit for future decision making.
2. To find optimum cost of media for advertising and to acquiring customers.
3. To get basis for treating food marts that are belongs to same category based sale, profit and cost

**Objectives:**

1. From EDA to get detail overview of factors that are affecting on sale as well as profit.
2. To get information about food products categories wise their sale and profit for different locations.
3. From past experience how much media impactful for sales and cost.
4. To make clusters of food marts based on how much they are similar with respect to sale profit and cost.
5. To build a machine learning model that predicts the optimum cost of media for advertisement.

**Tools Used:**

**Python**

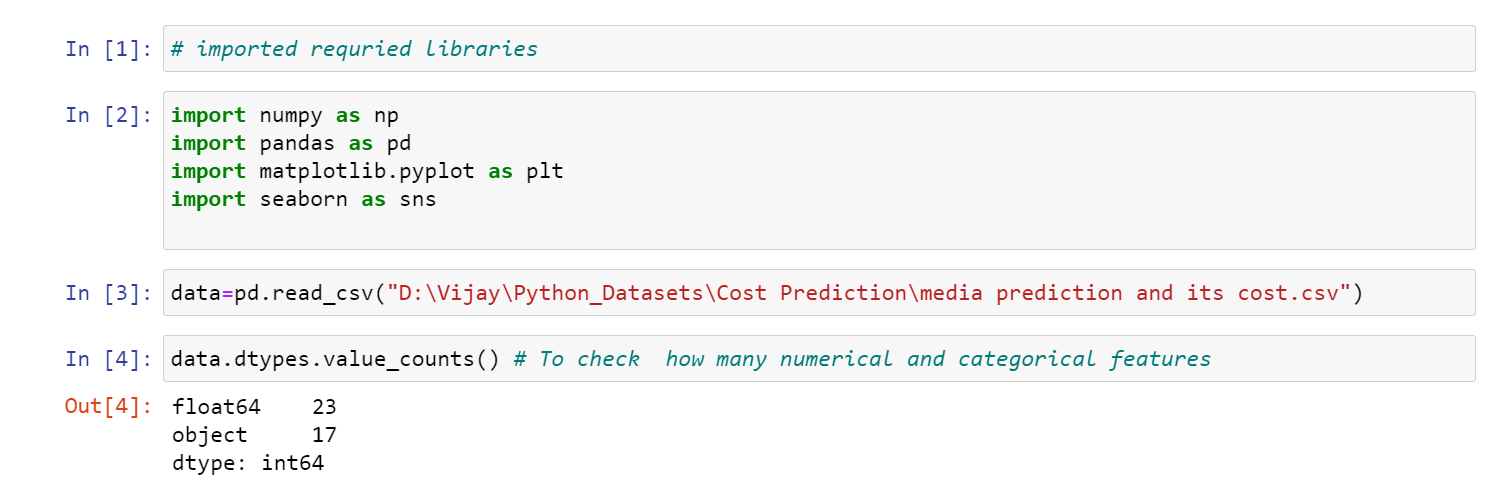
**MySQL**

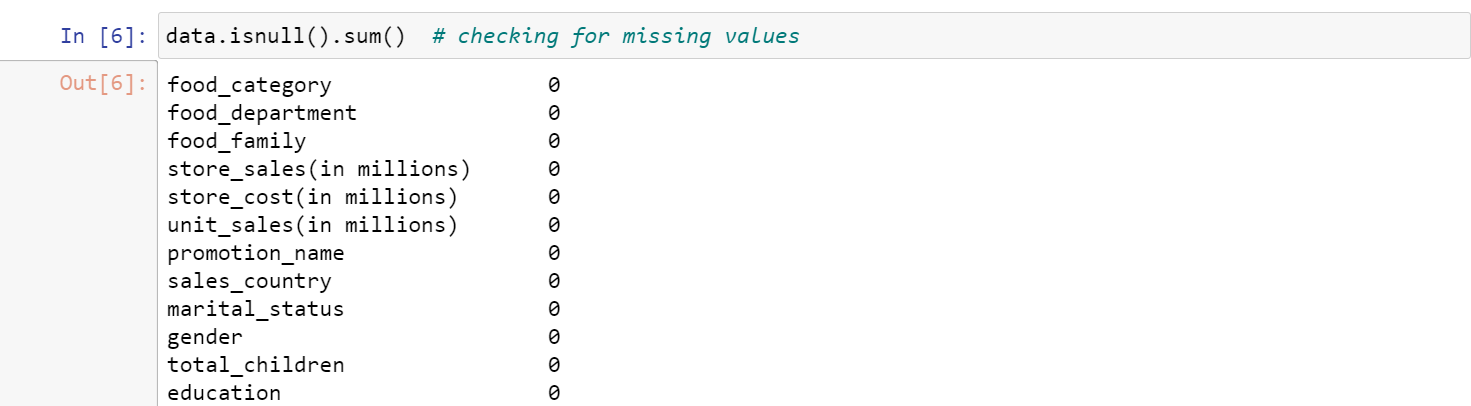
**Tableau**

**Hadoop**

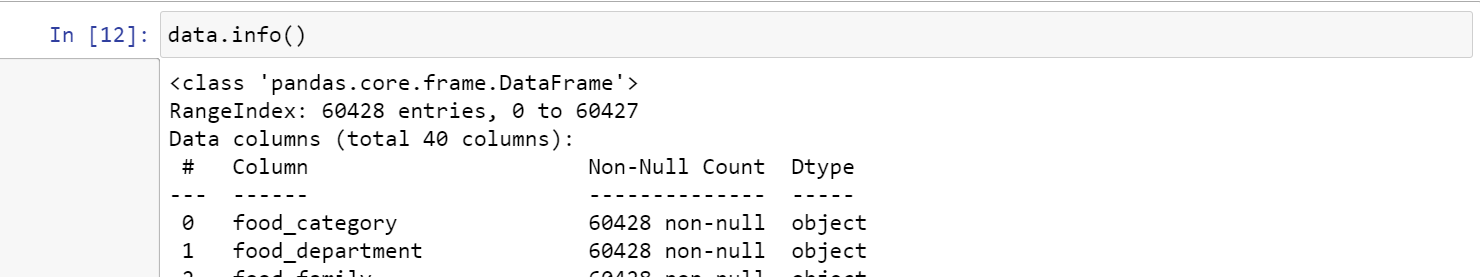
**Hive**

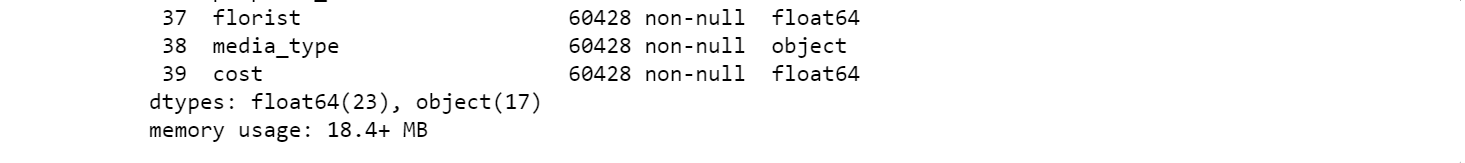
**Data Cleaning:**

****

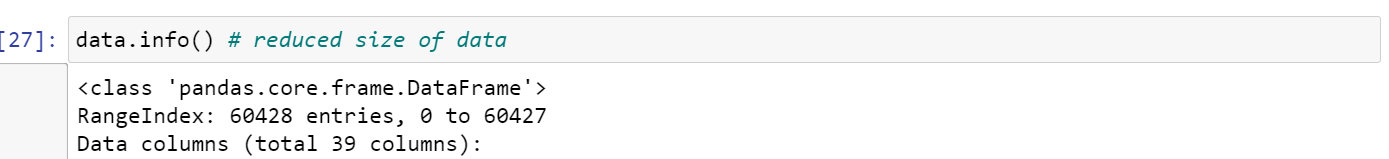
****

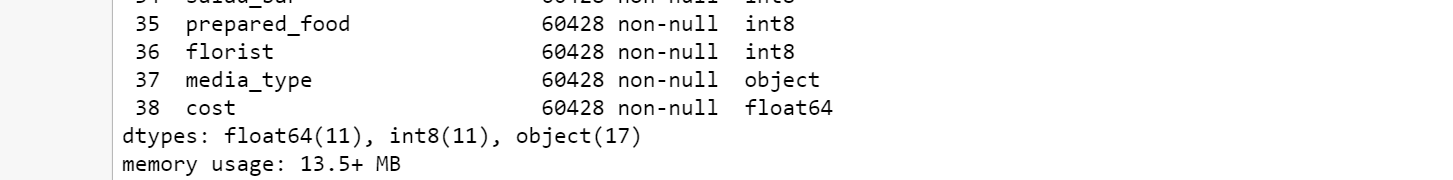
****

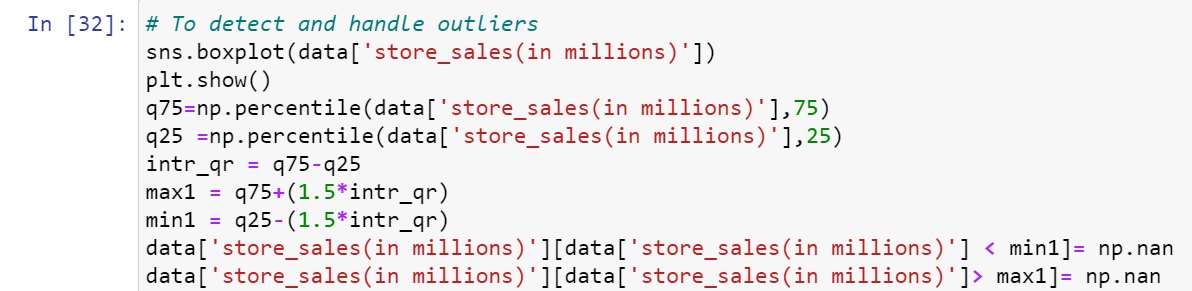
****

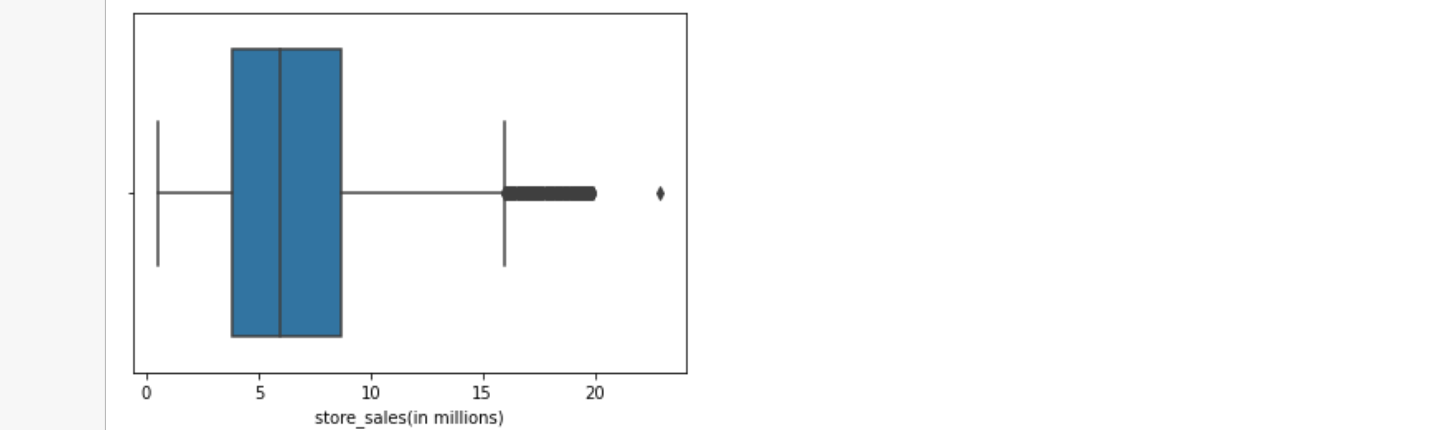
****

****

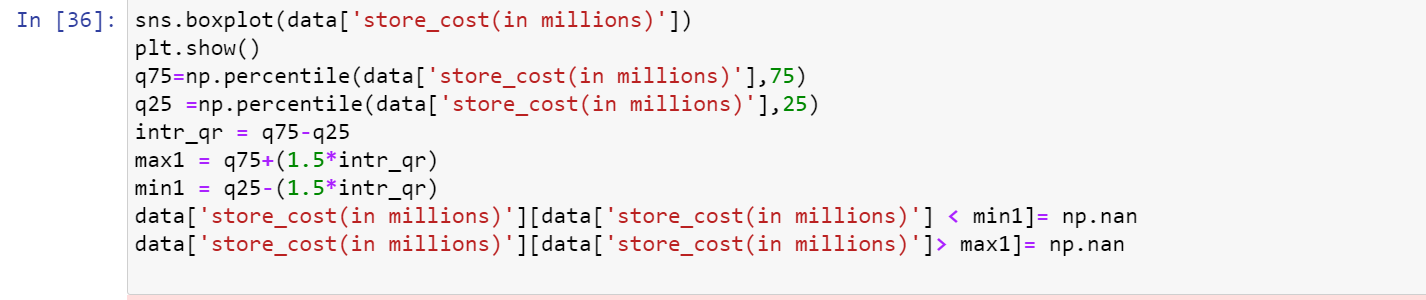
****

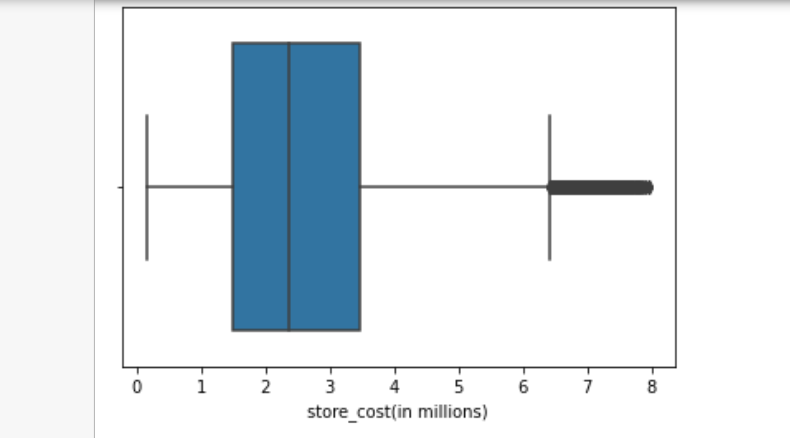
****

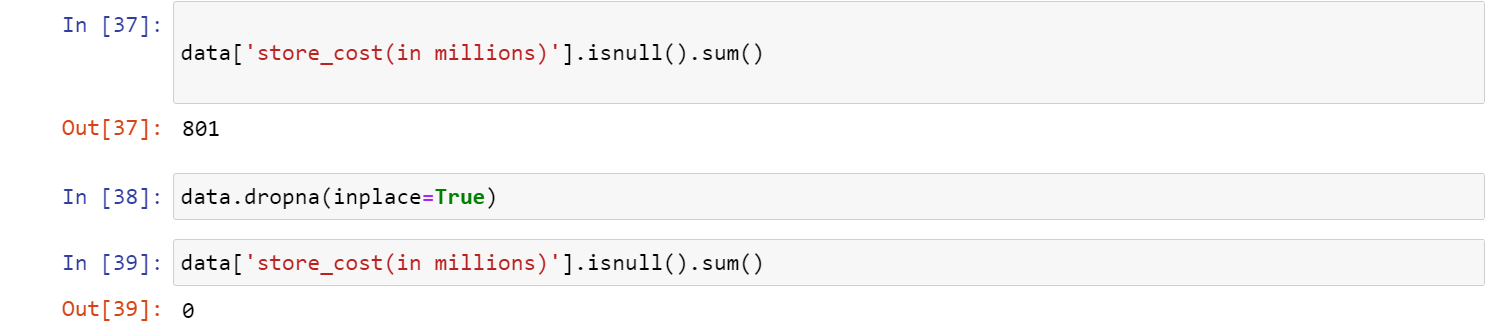
****

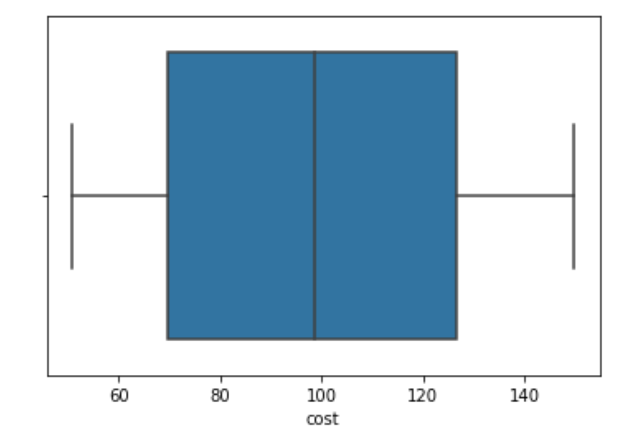
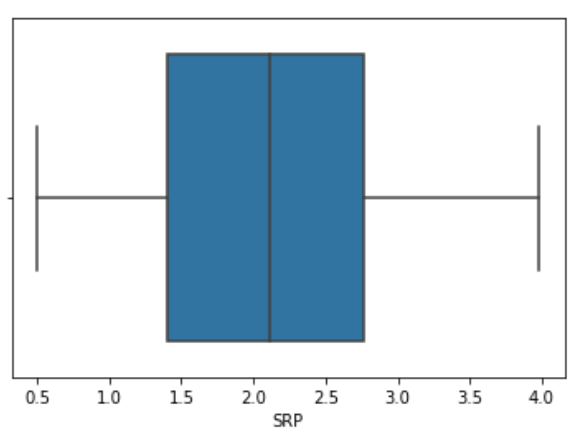
****

****

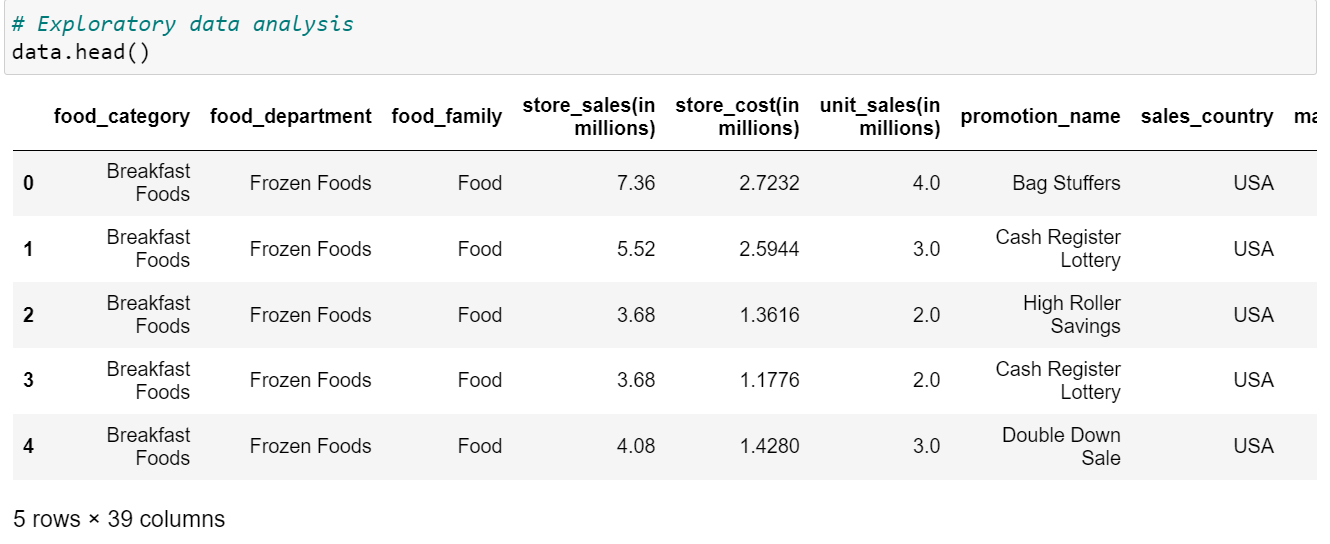
****

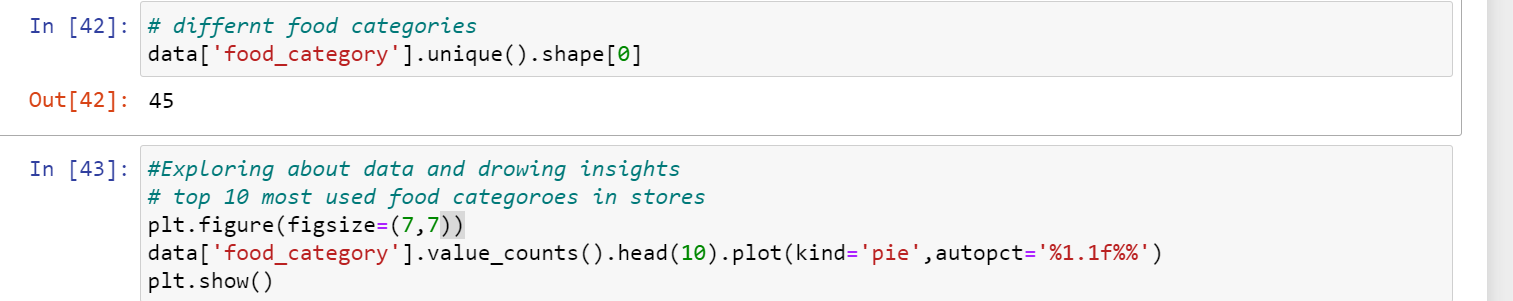
****

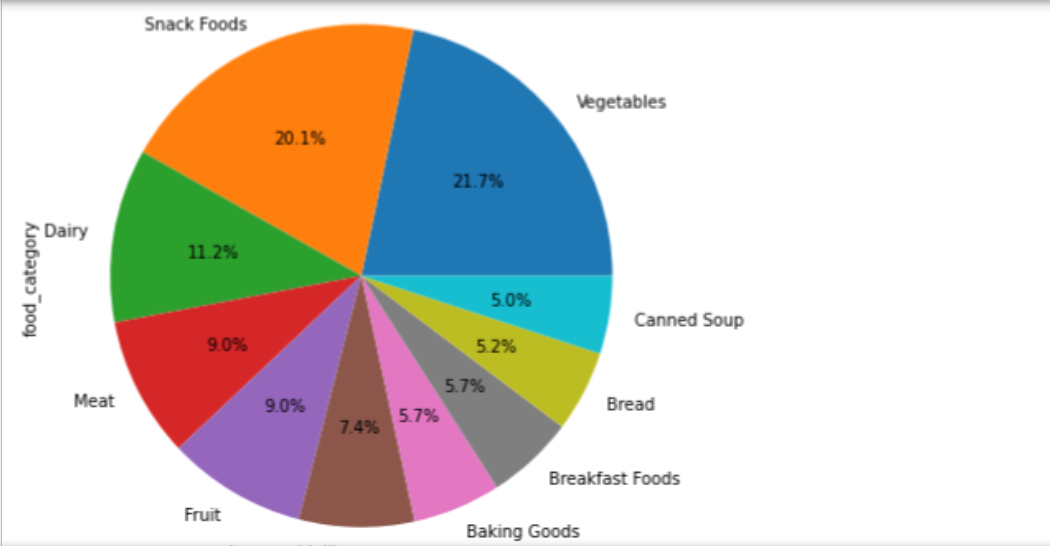
****

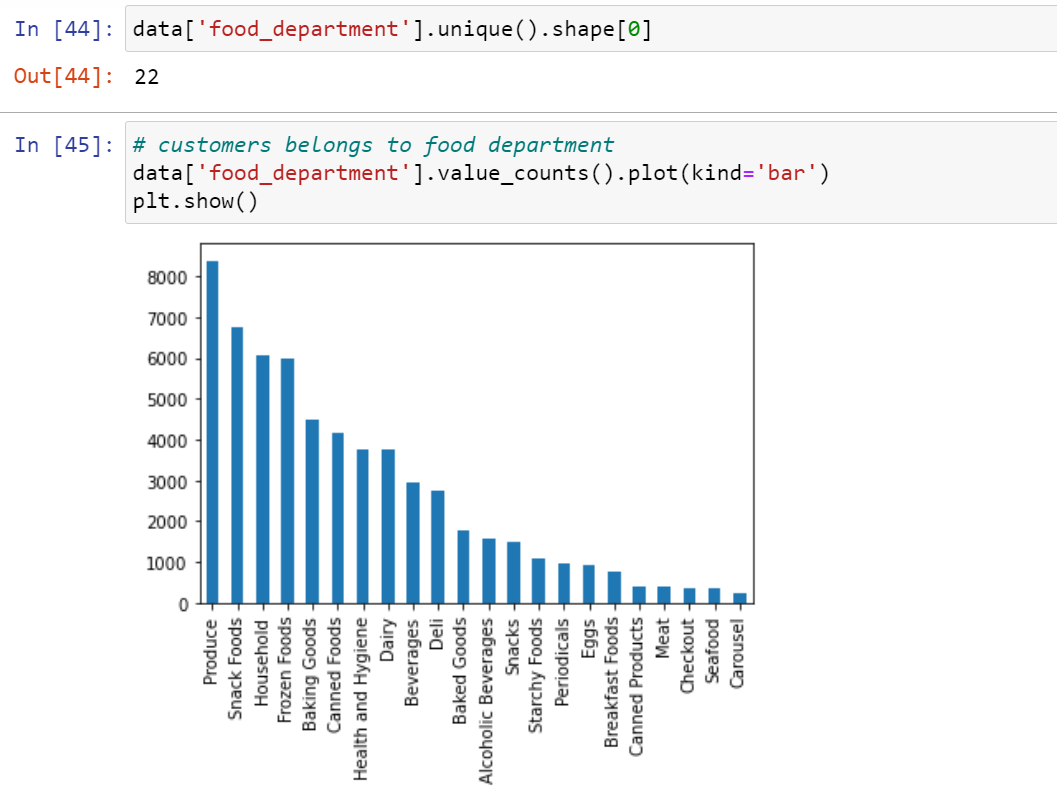
** **

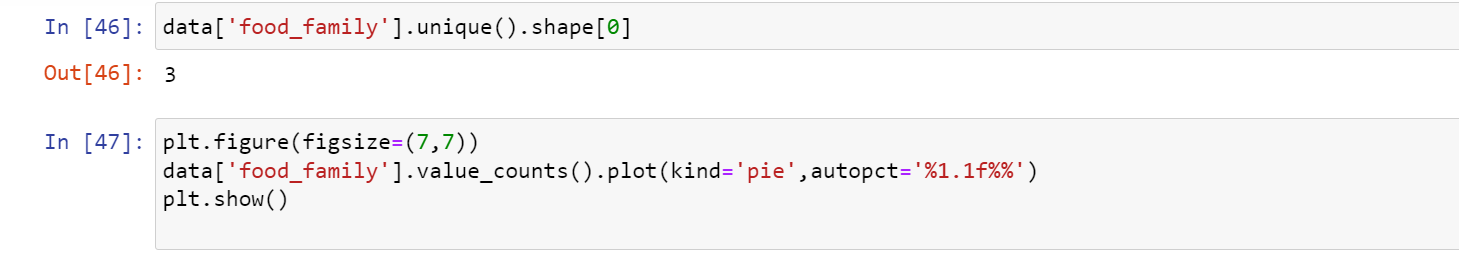
**Exploratory Data Analysis:**

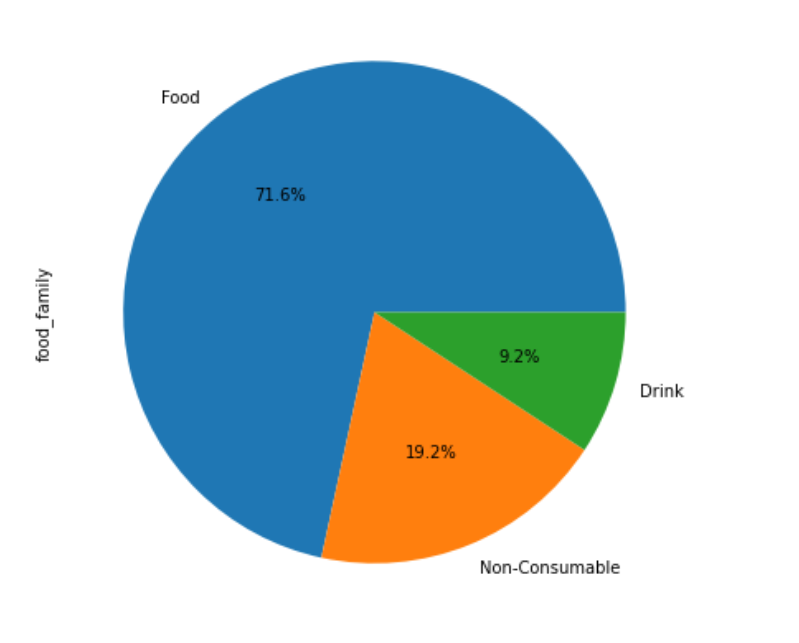


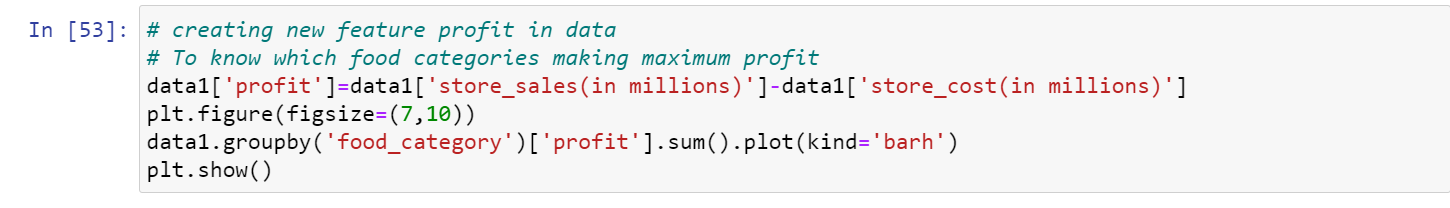
****

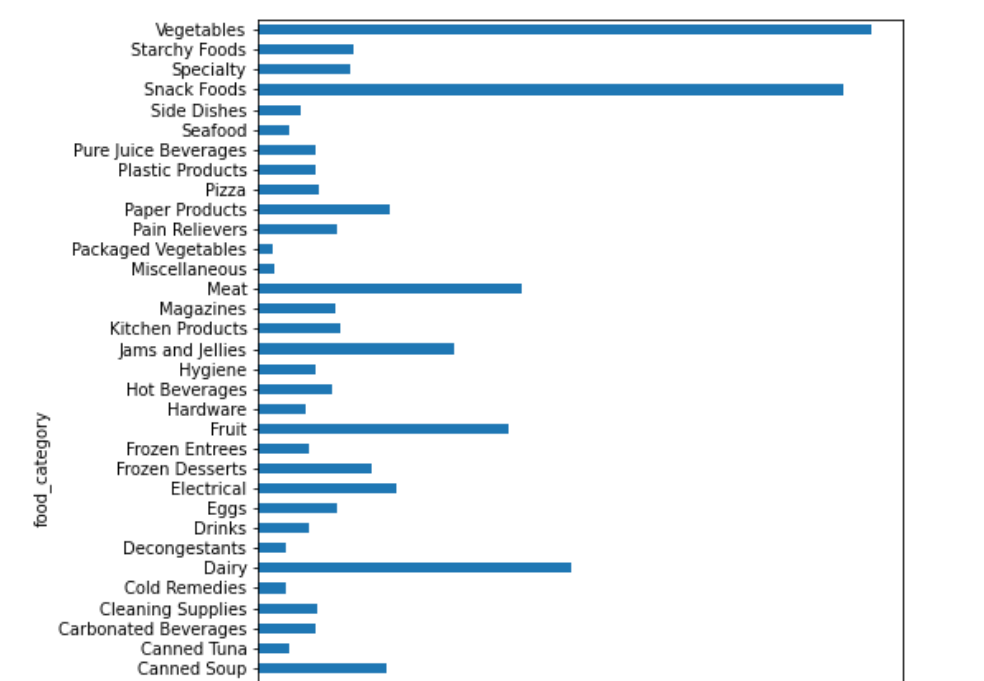
****

****

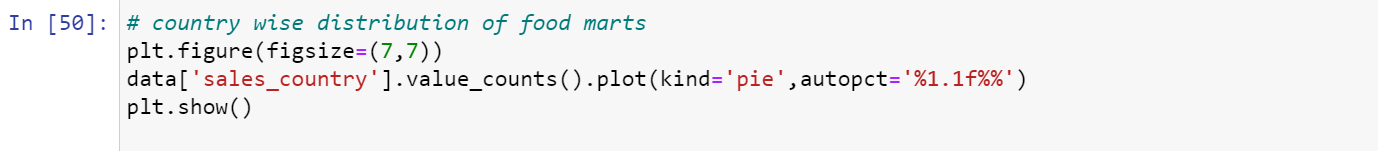
****

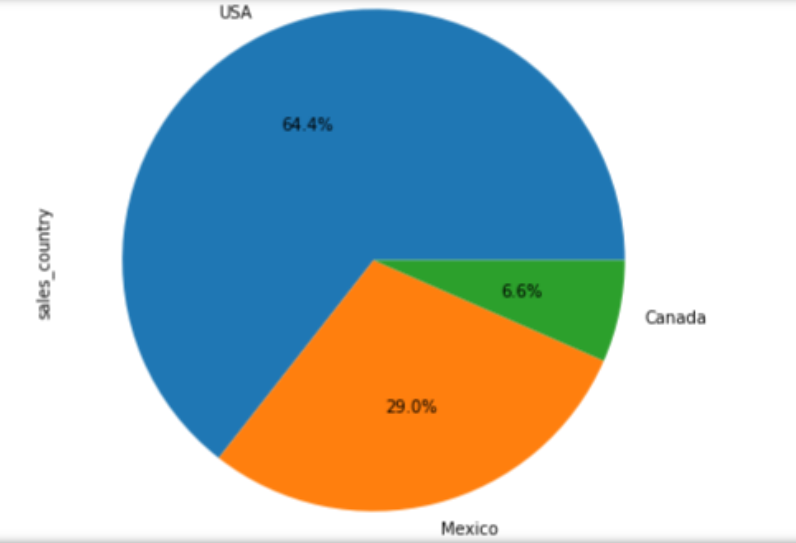
****

****

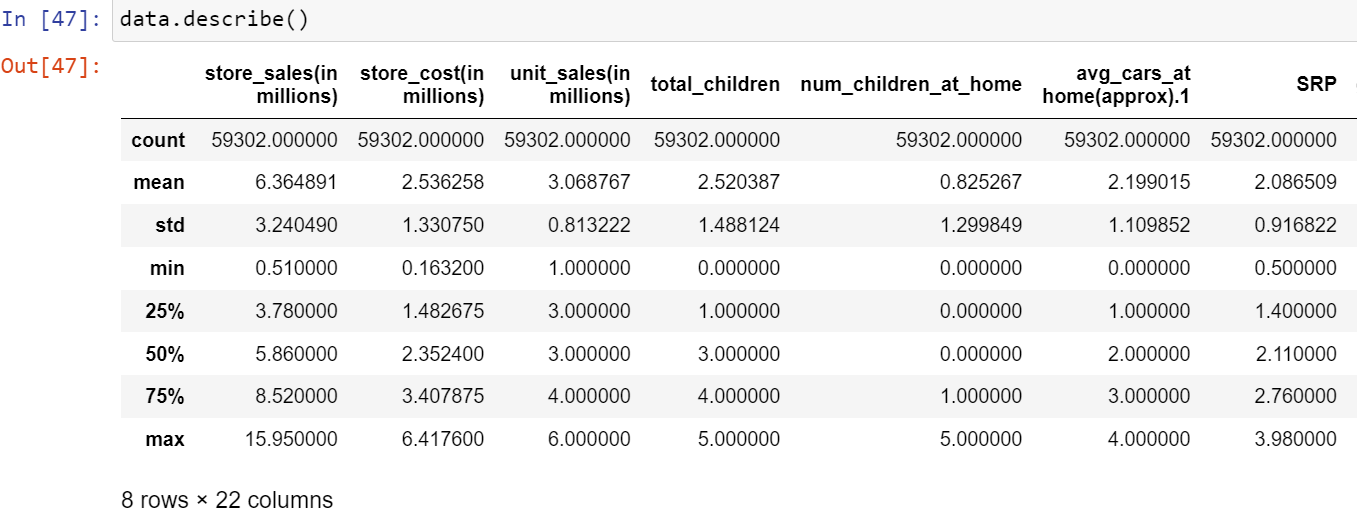
****

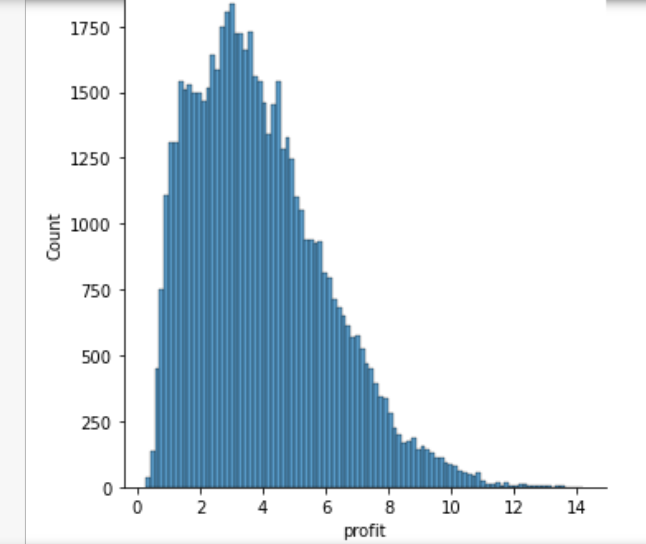
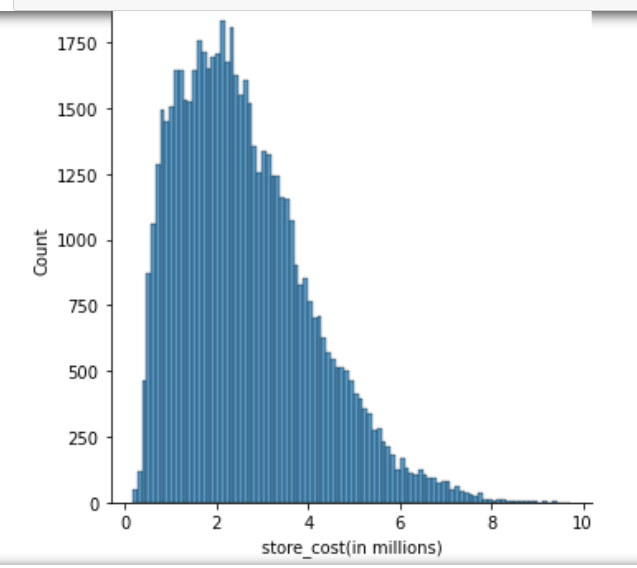
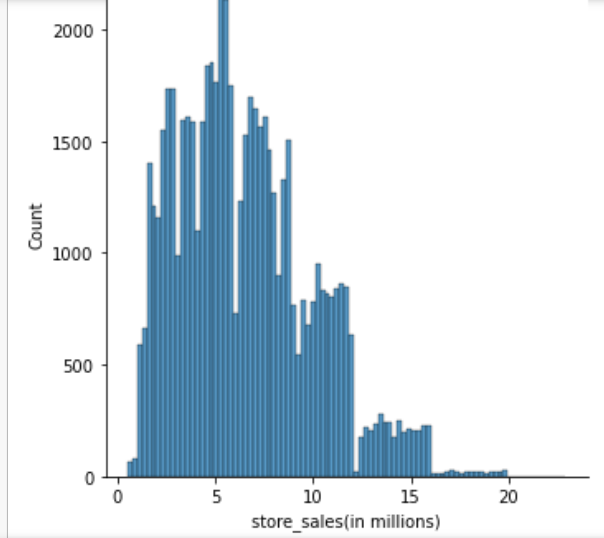
****

****

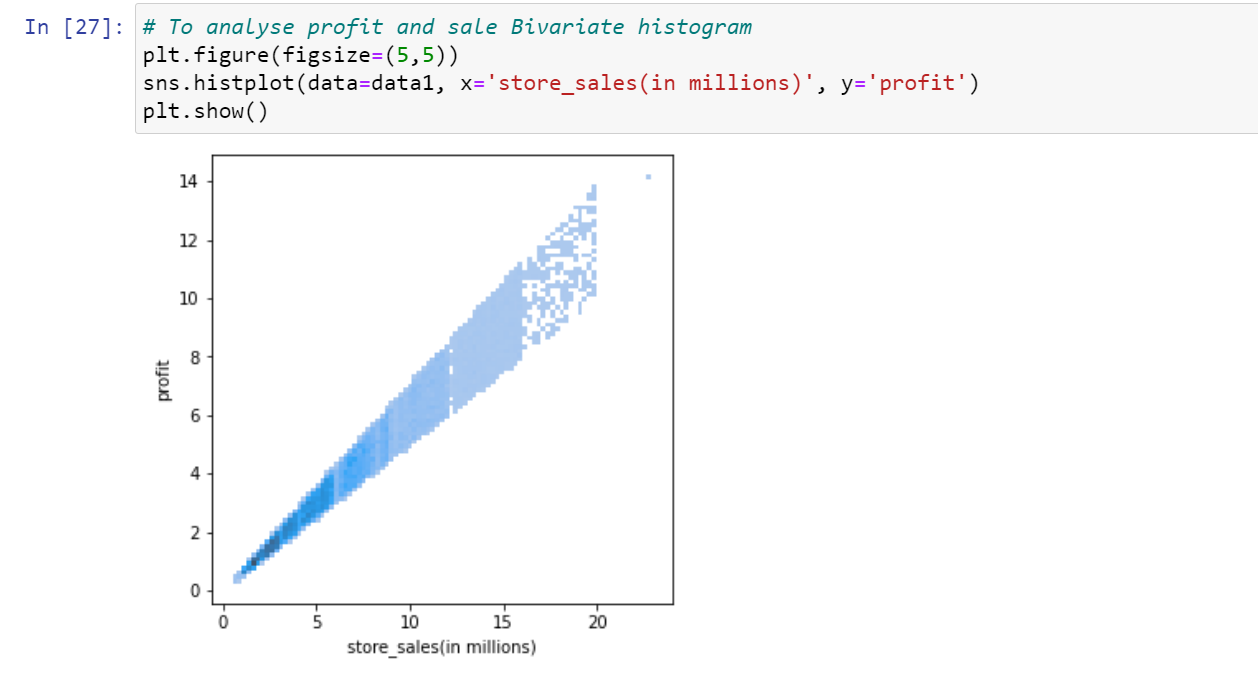
****

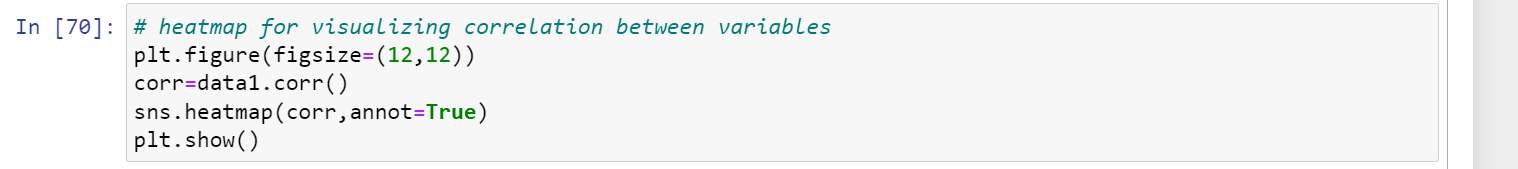
****

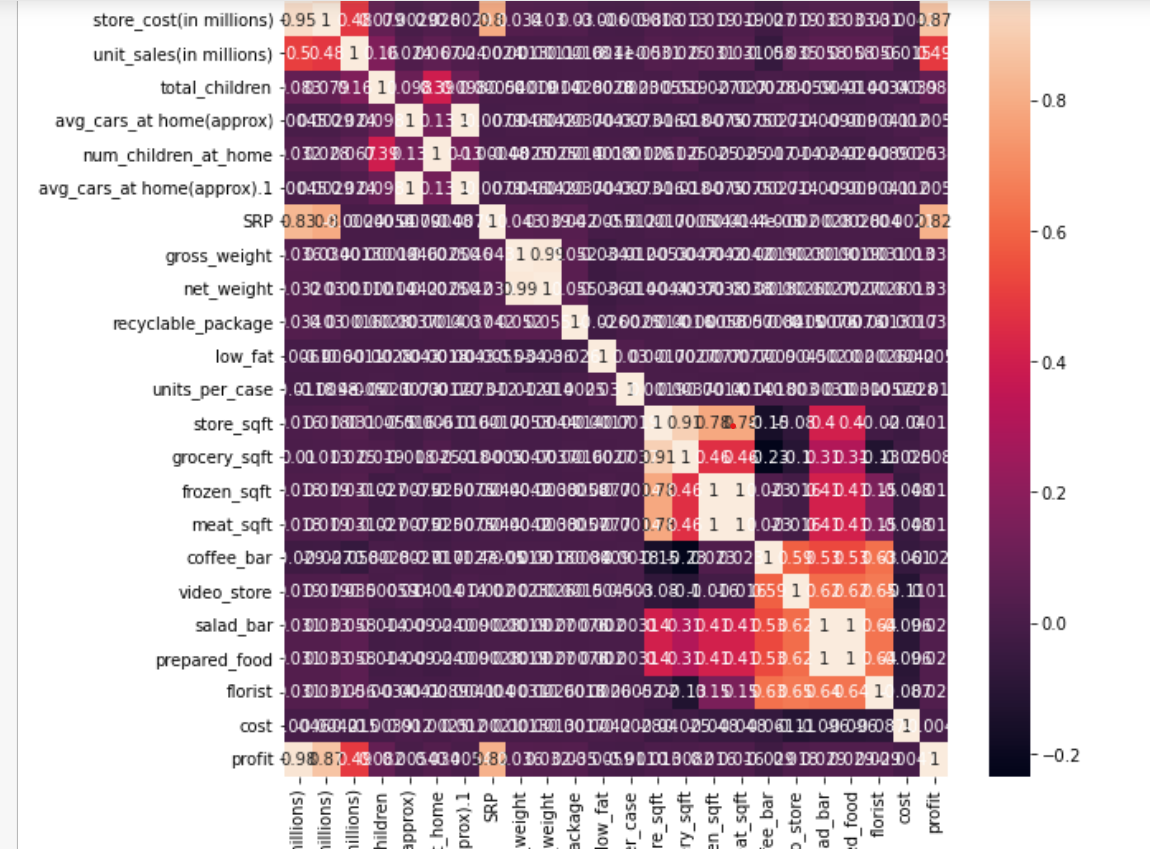
****

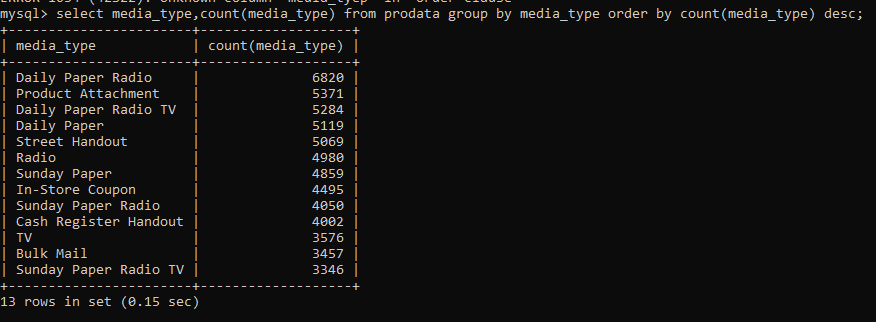
****

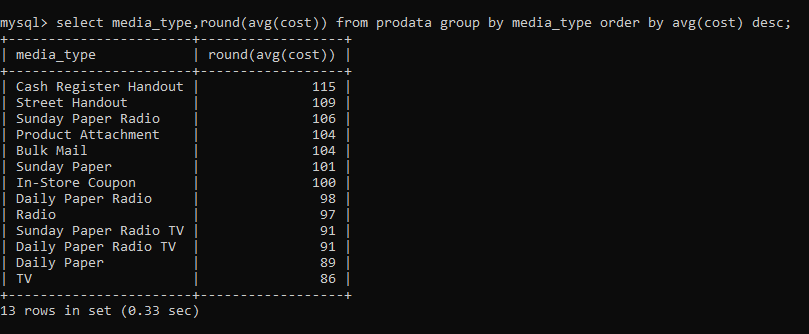
****

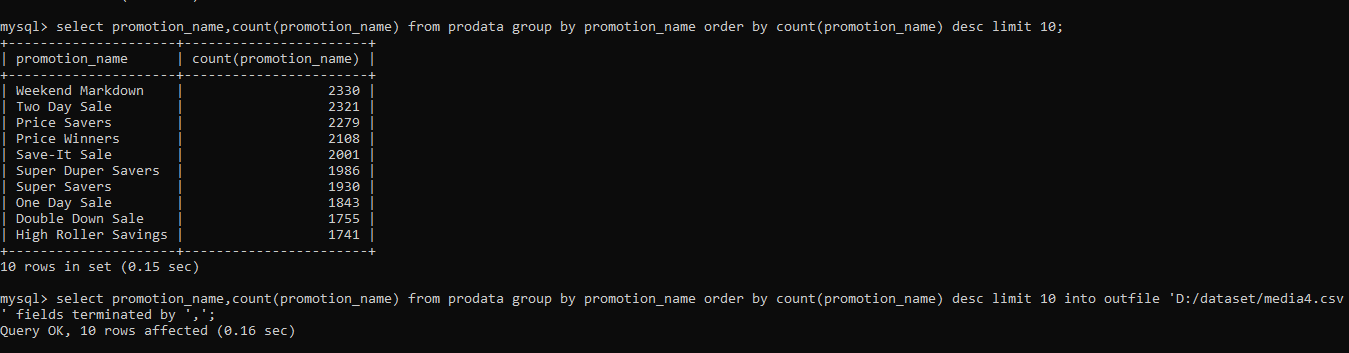
****

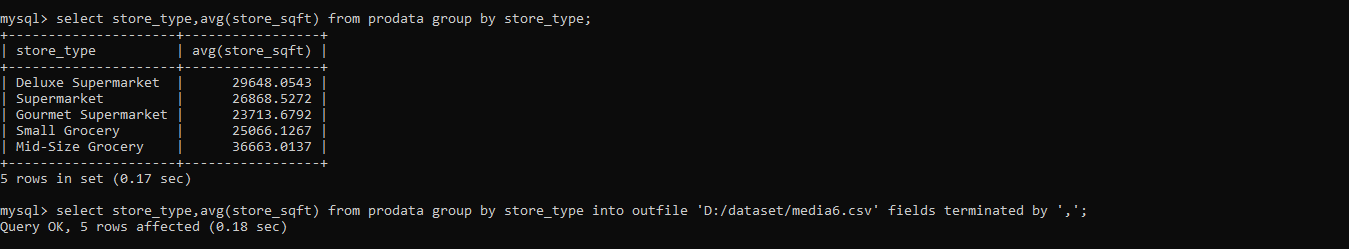




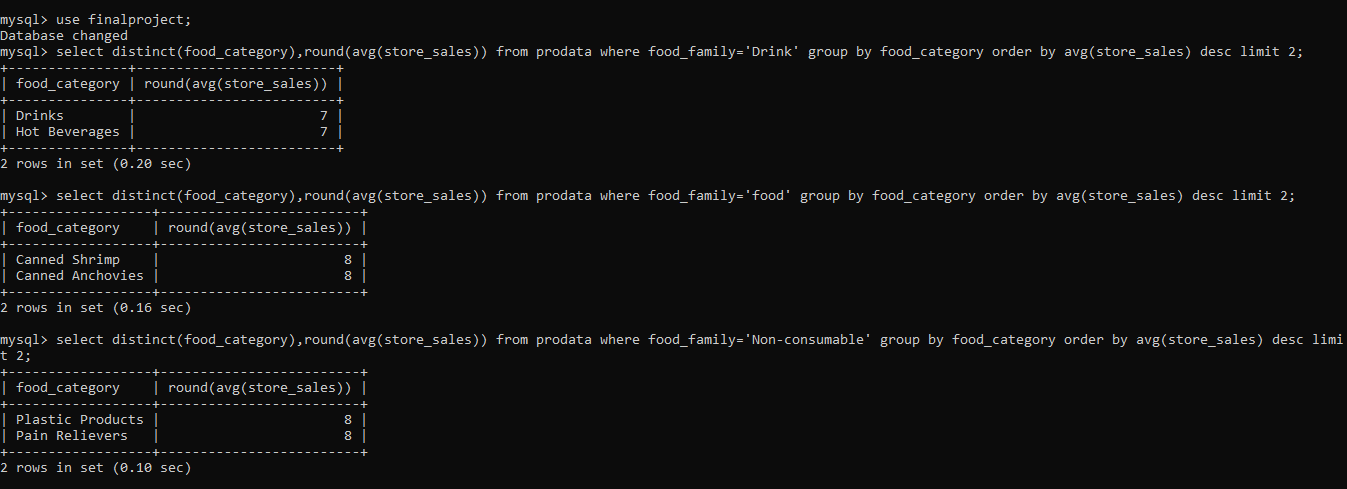




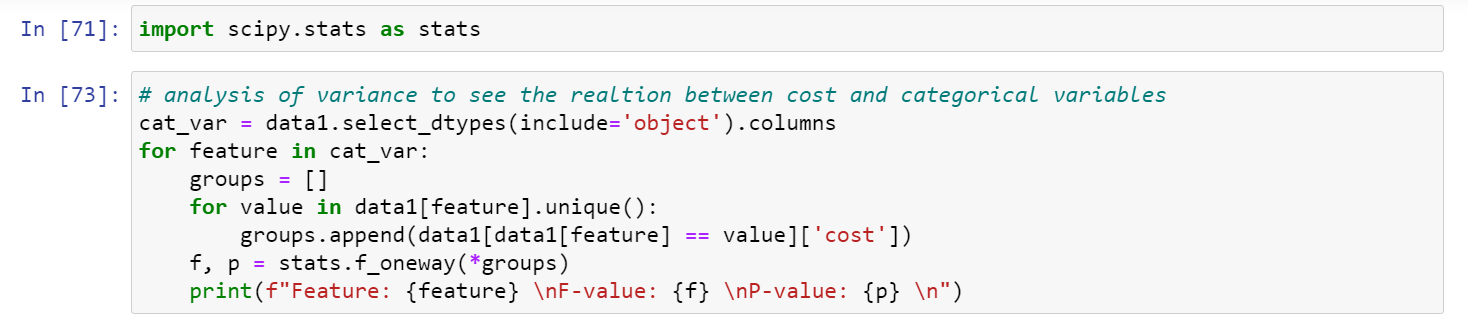


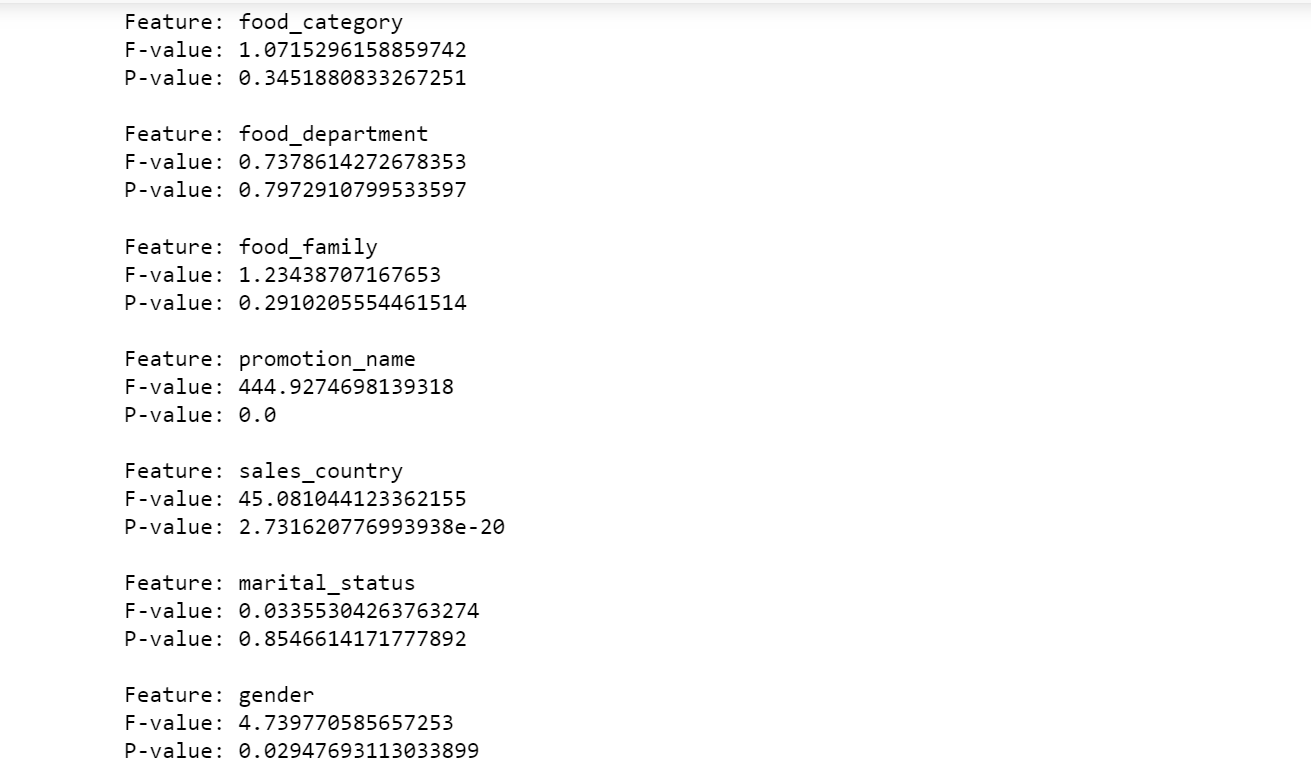


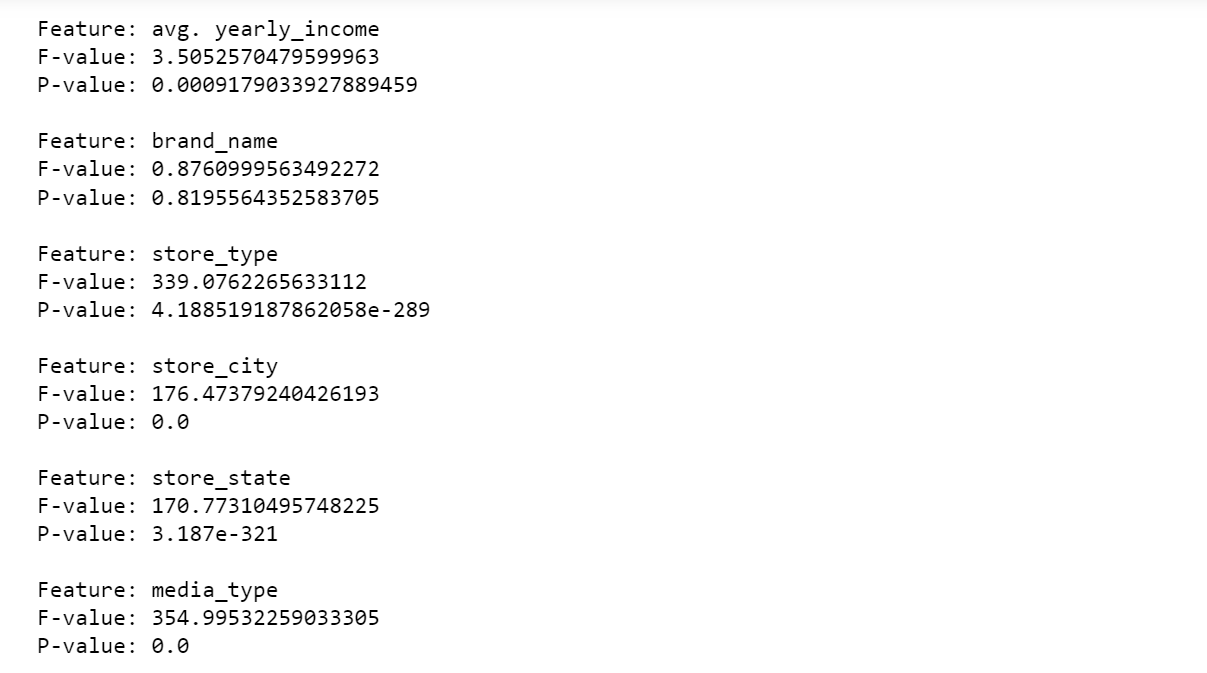
Top Two food categories for Drink, Food, Non-cousumable food families making maximum average sale.

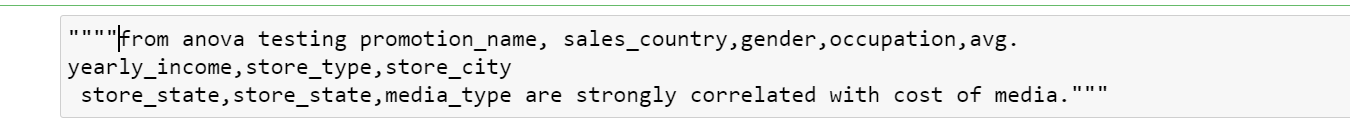


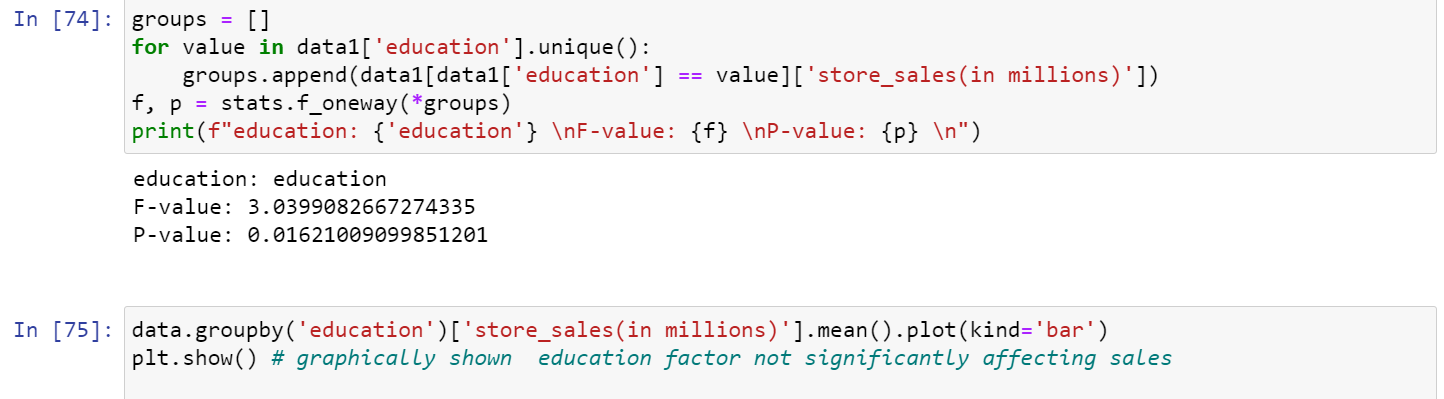
**Machine Learning:**

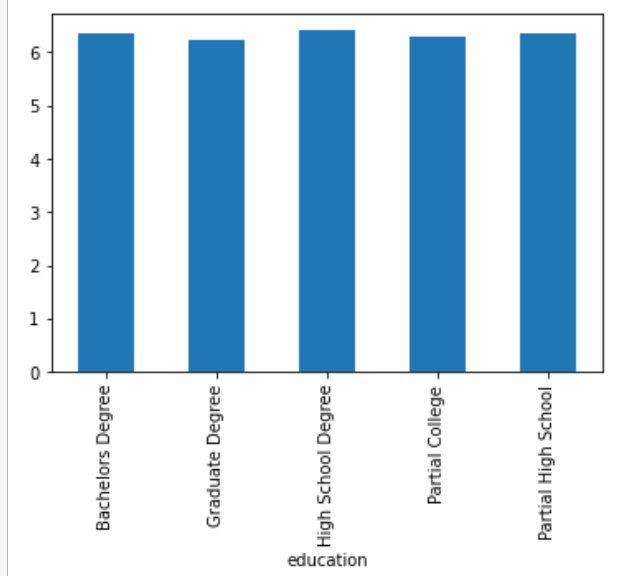
****

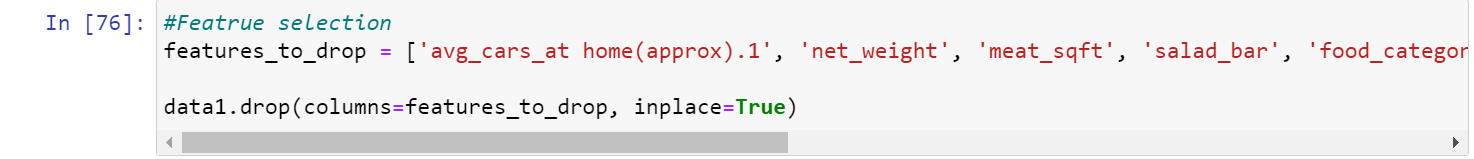
****

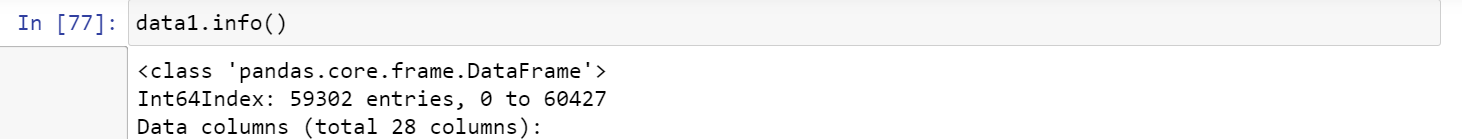
****

****

****

****

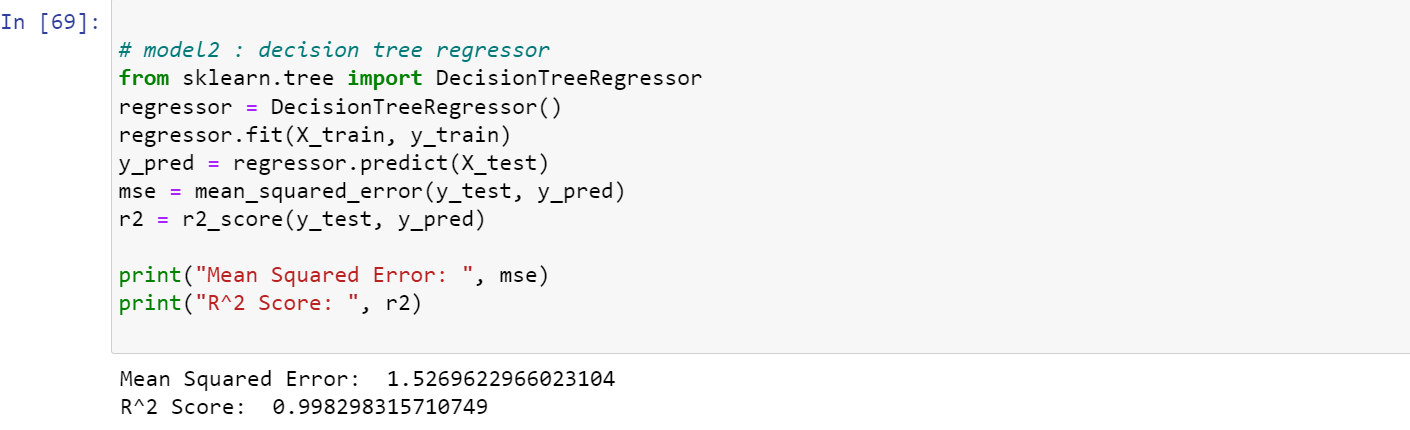
****

****

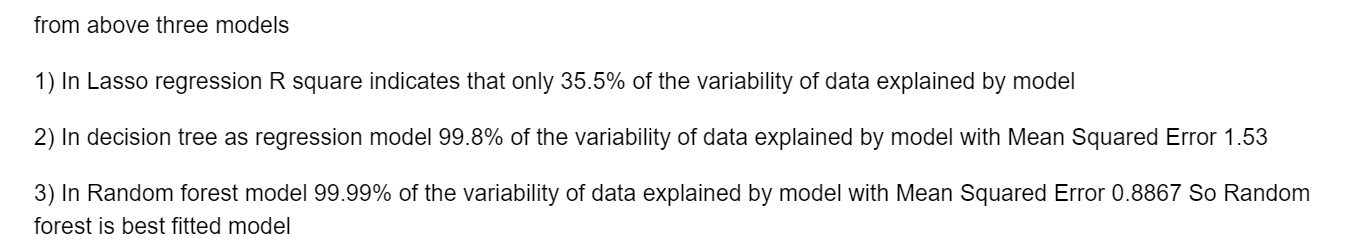
****

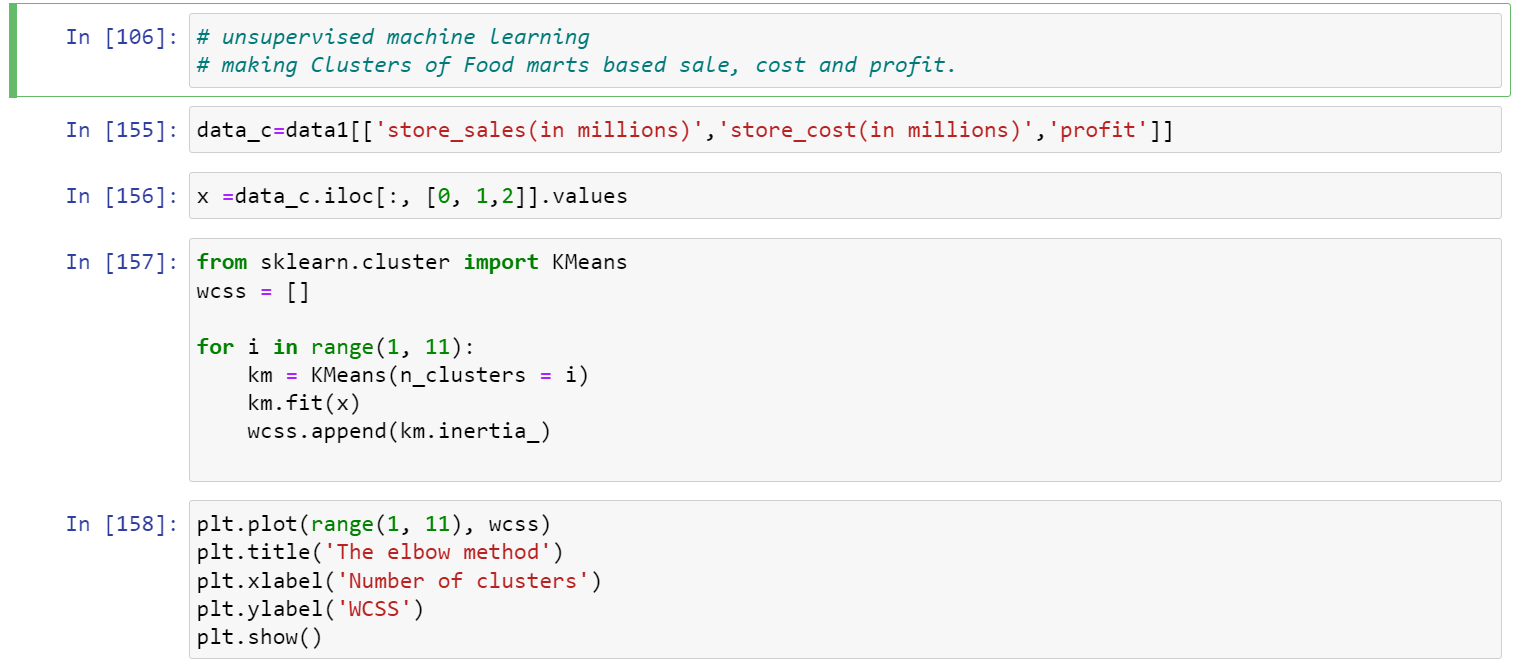
****

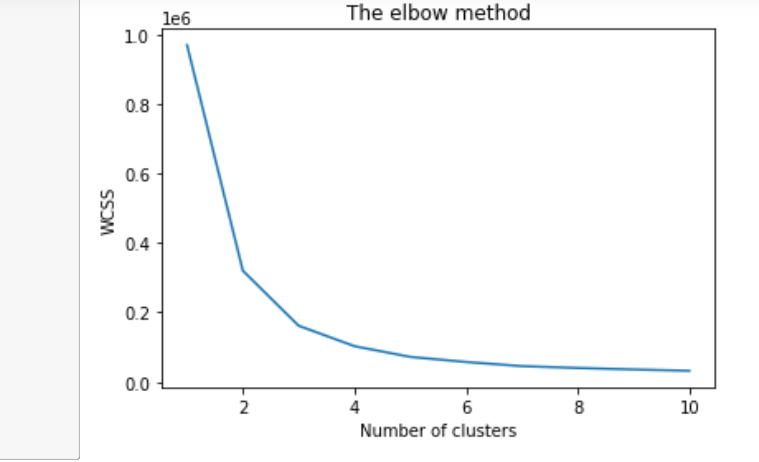
****

****

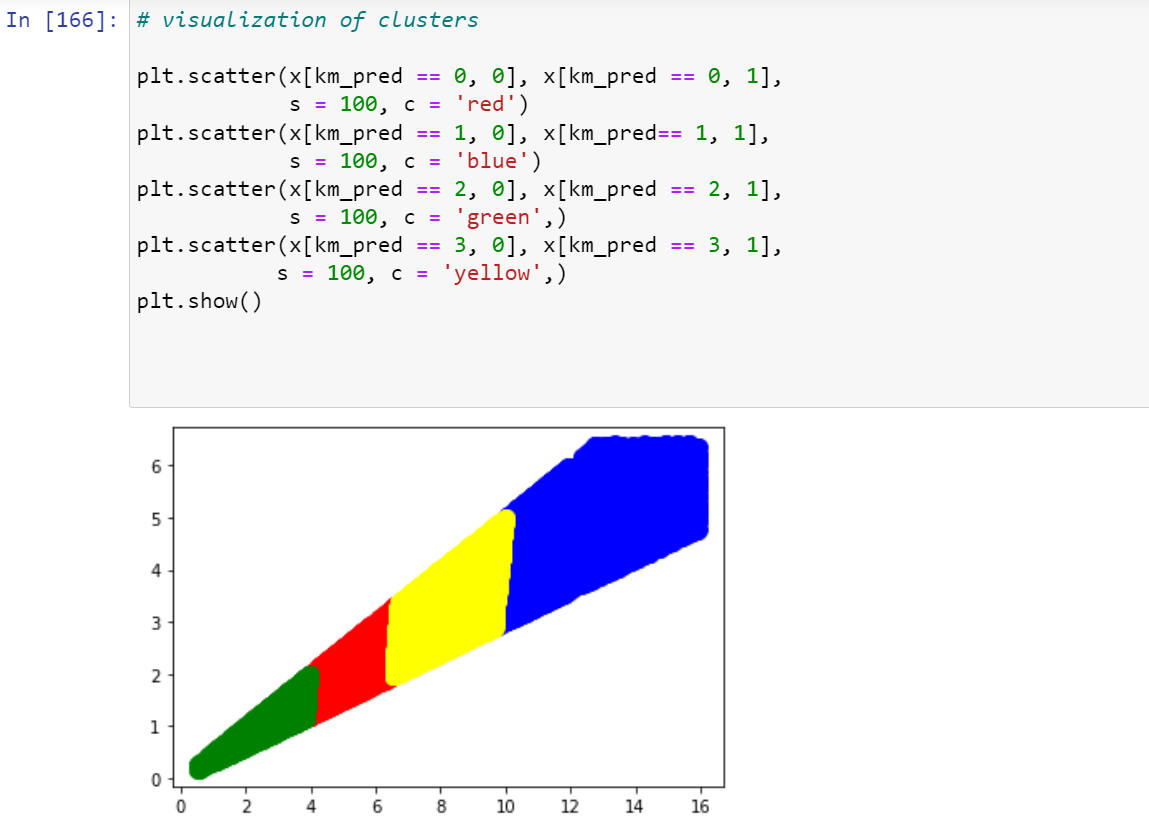
****

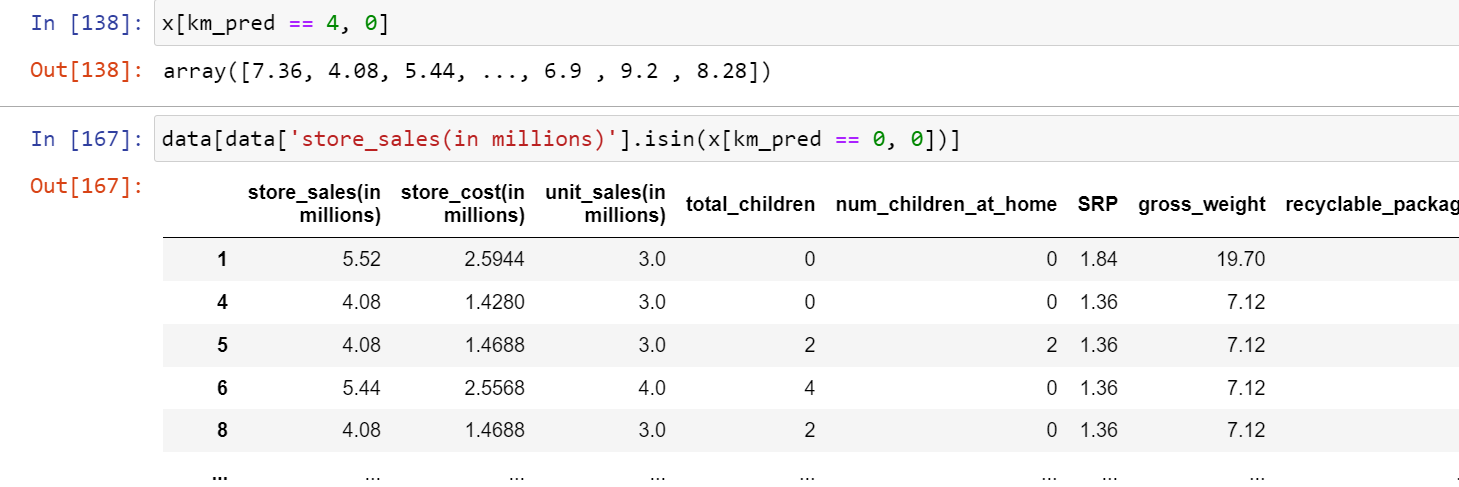
****

****

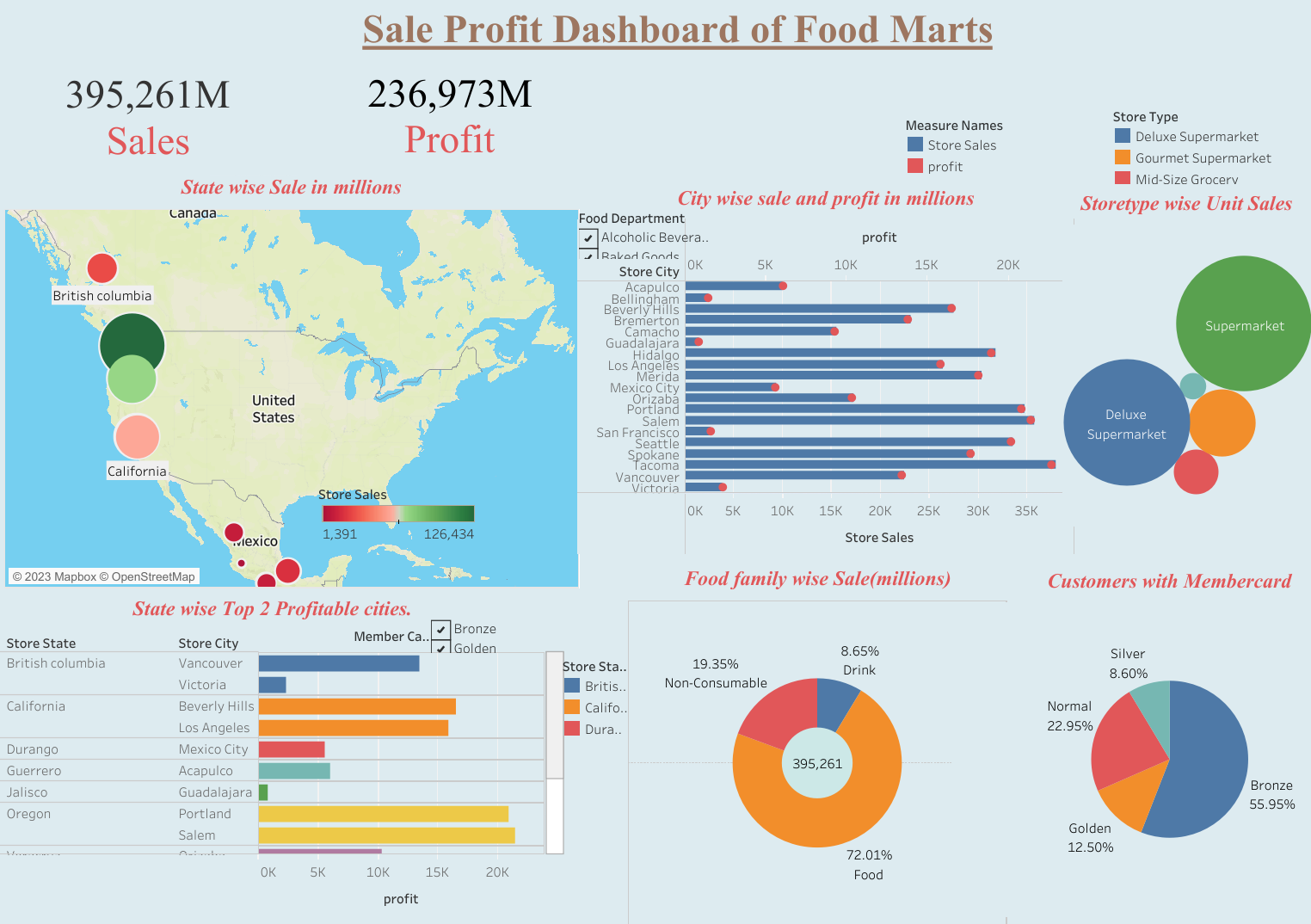
****

****

****

****

**Data Visualization:**



**Data Partitioning:**

There are two types of partitioning which is simple and dynamic here we are using dynamic partitioning

**Dynamic Partitioning**

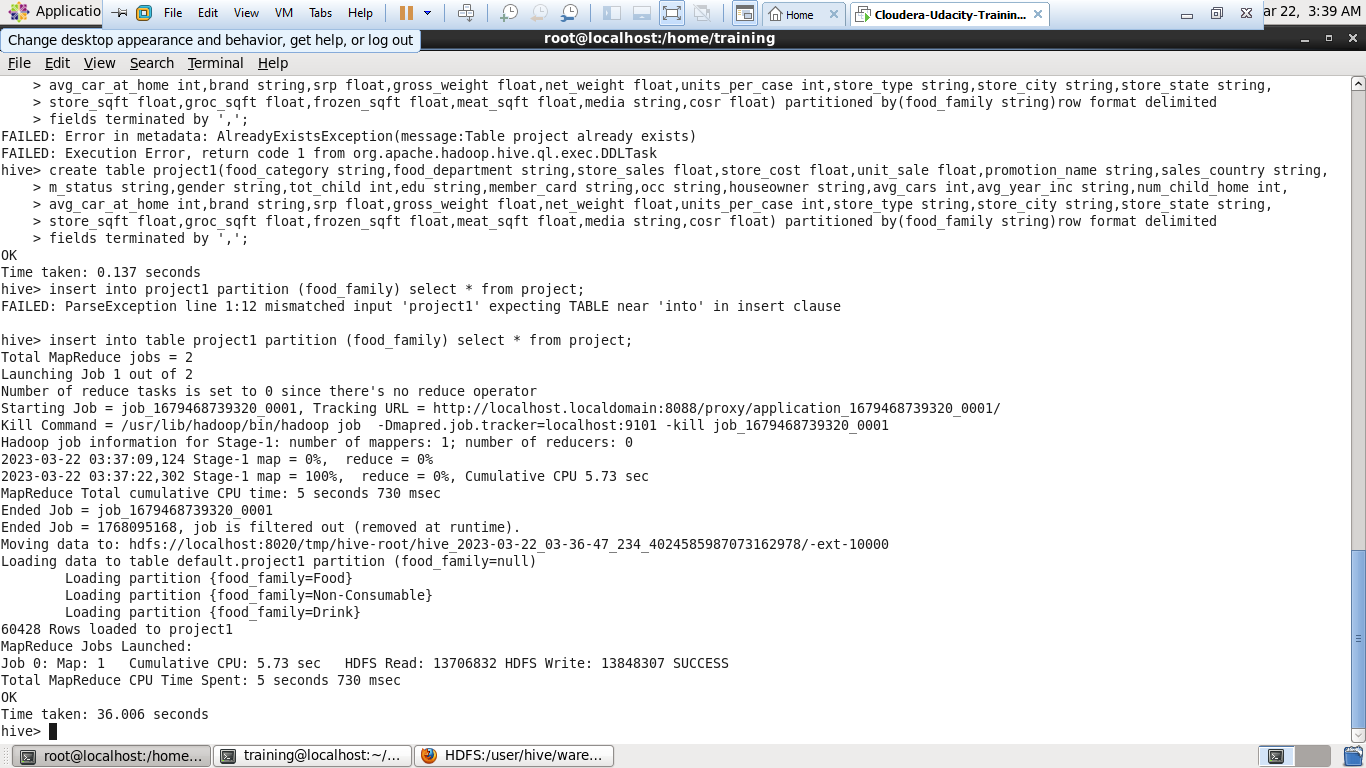
Single insert to partition table is known as the dynamic partition

To handle large amounts of data, dynamic partitioning is best choice

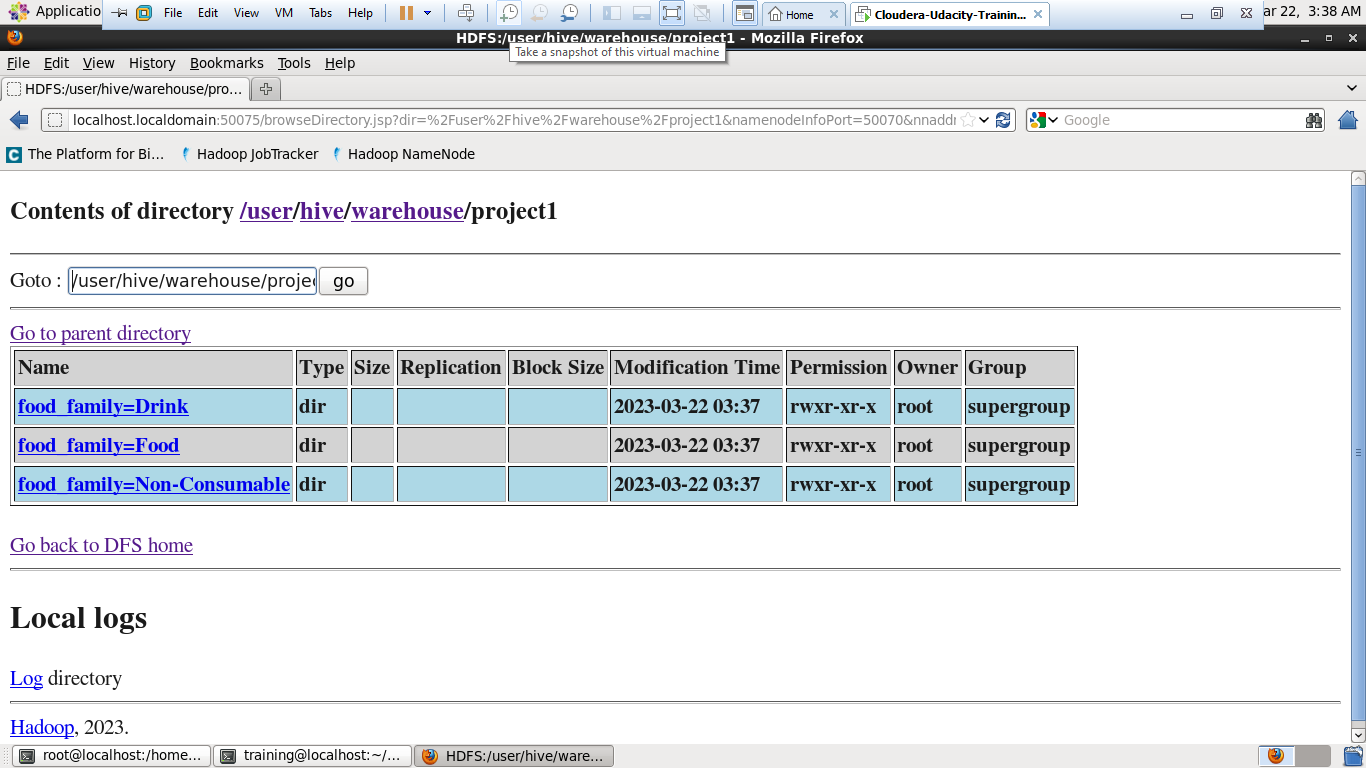
Dynamic partitioning can be performed on hive external and managed table

In dynamic partitioning there is no need of where clause

Create Dynamic Partitioned table:



**In Haddop NameNode we can see that Partitioned was happened on the basis of food family**



**Results:**

1. Vegetables, snack foods, dairy, meat and fruits are top 5 mostly sold food category.
2. Most of customers are from produce, snack foods and household food department.
3. Food family has maximum number of customers about 71.6% of total than Drink and Non-consumable food family.
4. Vegetables, snack foods and dairy products, making impactful sale while packed vegetables making low sale as well as profit as compared to others.
5. USA country has maximum no. of stores and Canada with lowest.
6. Salem and Tacoma are two cities which has maximum stores.
7. Maximum no. of stores are making sales up to 5 million and profit 3millions.
8. Most of stores using Daily paper and radio for advertising.
9. Cash register handout and street handout required maximum avg. cost for advertising.
10. Weekend markdown, Two-day sale and price savers are mostly used promotion
11. Mid-size grocery store has maximum avg. store size in sqft.
12. Food family drink Drinks and Hot beverages food categories making maximum average sale.
13. For food family canned Shrimp and Canned Anchovies machining maximum average sale.
14. For non-consumable Plastic products and Pain relievers making maximum average sales.
15. Random forest is best fitted model with mean squared error is 0.9470 and explains 99.89.
16. With K-Mean clusters made total 4 clusters of stores according to sale, cost and profit.
17. Tacoma, Salem and Portland cities making maximum sale as well as profit.
18. Food marts mode total 236923 M profit from sale of 395261 M.
19. Supermarkets and Deluxe-supermarkets are the store types where maximum unit are sold.
20. States Washington, Oregon, California from USA country made maximum sale with 126434M, 70440M, 56545M respectively.
21. About 55.95% of members having bronze member card.

**Conclusion:**

1. In this project we have explored more about the food marts data to give meaningful information to the franchise. So they can take decision about the business.
2. Apart from the results obtained above this project clould useful for food marts of USA, Canada, Mexico to make better decision of their media campaigns as well.
3. By understanding the different factors, the company can optimize their spending on promotions and advertising to effectively acquire new customers in food marts.
4. The supervised machine learning model will help them for predicting optimum value for cost of media campaigns for different food marts at different places.
5. The unsupervised model k-means will help them for treating the food marts based on which cluster they are and how much their sale, cost and profit.

This will help them for improve their weak areas in better way.