**MARKETING TARGETS**

A Course Project report submitted

in partial fulfillment of requirement for the award of degree

**BACHELOR OF TECHNOLOGY**

in

**ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING**

by

**B.SNEHITHA (2003A52021)**

**K.MAHALAXMI (2003A52034)**

**R.NAGARAJ (2003A52042)**

Under the guidance of

**Mr. NARESH KUMAR**

Assistant Professor, Department of CSE.

**

**Department of Computer Science and Artificial Intelligence**

**

**Department of Computer Science and Artificial Intelligence**

**CERTIFICATE**

This is to certify that project entitled **“MARKETING TARGETS**" is the bonafied work carried out by **B.SNEHITHA, MAHALAXMI, NAGARAJ** as a Course Project for the partial fulfillment to award the degree **BACHELOR OF TECHNOLOGY** in **ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING** during the academic year 2022-2023 under our guidance and Supervision.

**Mr. S Naresh Kumar Dr. M.Sheshikala**

Asst. Professor, Assoc. Prof. & HOD (CSE),

S R University, S R University,

Ananthasagar, Warangal. Ananthasagar, Warangal.

**ACKNOWLEDGEMEN**

We express our thanks to Course co-coordinator **Mr.** **S.Naresh Kumar, Asst. Prof.** for guiding us from the beginning through the end of the Course Project. We express our gratitude toHead of the department CS&AI, **Dr.M.Sheshikala, Associate Professor** for encouragement, support and insightful suggestions. We truly value their consistent feedback on our progress, which was always constructive and encouraging and ultimately drove us to the right direction.

We wish to take this opportunity to express our sincere gratitude and deep sense of respect to our beloved Dean, School of Computer Science and Artificial Intelligence, **Dr C. V. Guru Rao**, for his continuous support and guidance to complete this project in the institute.

Finally, we express our thanks to all the teaching and non-teaching staff of the department for their suggestions and timely support.

**ABSTRACT**

Marketing data analysis is a technique where the business will take all the available information regarding the market and come up with a marketing plan. It is a very vital piece of activity for any sort of business. It also shows you how well you have done in the market using your current marketing techniques. Already existing solutions would be the predictive sales after marketing but in this problem statement we used various modes of marketing. The future scope for the project will be that as the main aim of this project is to know the relation among the sales ang the mode of marketing platform, the users can analyze and choose best platform so that they get maximum profits. By using this they need not invest their money in all the modes of marketing but instead choose a best one. In other mays it will also help the user to know how much average amount need to be spent for marketing.

**Table of Contents**

**Chapter No. Title Page No.**

1. Introduction
   1. Overview 1
   2. Problem Statement 1
   3. Existing system 1
   4. Proposed system 1
   5. Objectives 2
   6. Architecture 2
2. Literature survey
   * 1. Document the survey done by you 3
3. Data pre-processing
   1. Dataset description 5
   2. Data cleaning 6
   3. Data augmentation 7
   4. Data Visualization 8
4. Methodology
   1. Procedure to solve the given problem 12
   2. Model architecture 13
   3. Software description 15
5. Results and discussion 16
6. Conclusion and future scope 19
7. References 20

**LIST OF FIGURES**

|  |  |  |
| --- | --- | --- |
| **S.N0** | **NAME** | **PG.NO** |
| 1 | Architecture | 2 |
| 2 | Model Architecture | 12 |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |

**LIST OF TABLES**

|  |  |  |
| --- | --- | --- |
| **S.N0** | **NAME** | **PG.NO** |
| 1 | Describing dataset | 4 |
|  |  |  |
|  |  |  |

**1.INTRODUCTION**

#### 1.1 OVERVIEW

Product advertising is the organizational strategy of increasing sales by promoting products to potential customers.

**Product advertising** is any method of communication about the promotion of a product in an attempt to induce potential customers to purchase the product. Advertisement usually requires payment to a communication channel. The general objective is to increase brand awareness or to demonstrate the differences between product and competing products in order to sell them.

Product advertisement can be done through numerous communication channels. Some common channels are listed here:

* Television
* Radio
* Print (newspapers and magazines)
* Mail (flyers, circulars, coupons)
* Websites
* Social media
* Signs and billboard

#### 1.2 PROBLEM STATEMENT

#### The problem statement we encountered is to find out the difference between the previous sales as well as the sales once after the product is being advertised through different platforms.

#### 1.3 EXISTING SYSTEM

Here we created our own dataset by collecting the information if 10 different brands and their investment for advertising their product in different platforms for the years 2000-10.

#### 1.4 PROPOSED SYSTEM

The Components we used to train our model are:

Google collab,

Collab notebooks execute code on Google's cloud servers.

Page | 1

#### 1.5 DEFINE OBJECTIVES

Objective of this paper is to develop a system that predicts the rate of sales in specific year is increased or decreased. This is taken into consideration after observing all the amount spent on different modes of advertising like tv ads, radio ad, social media sponsoring and being advertised or suggested by any of the influencer.

#### 1.6 OVERALL ARCHITECTURE

Data Collection

Preparing the

Data

C

hoosing a

model

Training the

Model

Evaluating the

Model

P

arameter

Tuning

Making

Predictions

Page | 2

**2.LITERATURE SURVEY**

##### **2.1 SURVEY:**

Once after selecting the problem statement the first thing, we started with is the collection if the dataset. In which we took the help of bowser to know how on an average a brand or company spends for advertising their product. Later after collecting all the data in different attributes, we collected the values of their previous sales through browsing networks.

##### **2.2 RELATED WORK**

1. **PYTHON -** Python is an interpreted, high-level, general-purpose programming language.

Python's design philosophy emphasizes code readability with its notable use of significant whitespace. Its language constructs and object-oriented approach aim to help programmers write clear, logical code for small and large-scale projects. Python is dynamically typed AND supports multiple programming paradigms, including procedural, object-oriented, and functional programming.

1. **JUPYTER LAB -** Jupyter lab is the latest web-based interactive development environment for notebooks, code, and data. Its flexible interface allows users to configure and arrange workflows in data science, scientific computing, computational journalism, and machine learning. A modular design invites extensions to expand and enrich functionality.

Page | 3

**3.DATA PREPROCESSING**

##### **3.1 DESCRIBING DATASET :**

The dataset we created consists of 350 rows and 10 columns i.e.., 10 attributes.

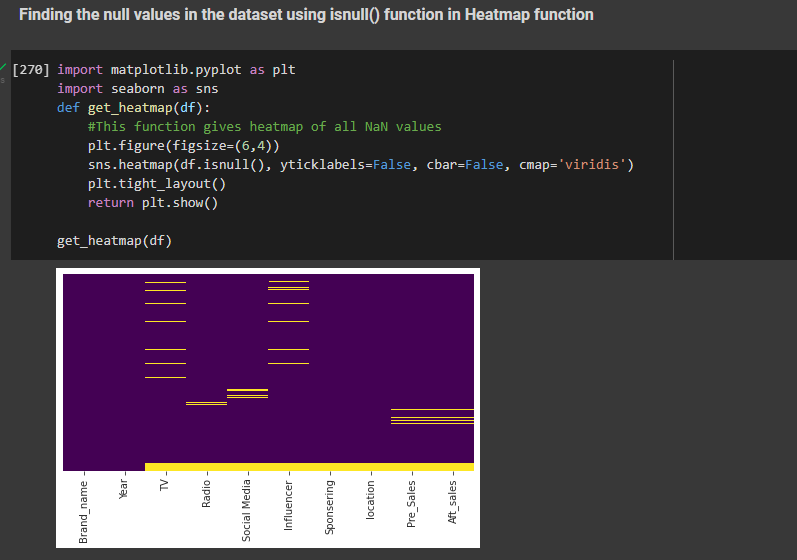
In detail,

|  |  |  |
| --- | --- | --- |
| **S.NO** | **ATTRIBUTE** | **DESCRIPTION** |
| 1 | Brand\_name | Name of the Brands(Companies) |
| 2 | Year | Year between 2000-10 |
| 3 | TV | Amount spend for advertising in TV in millions |
| 4 | Radio | Amount spend for advertising in Radio in millions |
| 5 | Social Media | Amount spend for advertising in Social Media in millions |
| 6 | Influencer | Amount spend for advertising in Influencer in millions |
| 7 | Sponsoring | Amount spend for advertising in Sponsoring in millions |
| 8 | location | Urban or Rural |
| 9 | Pre\_Sales | Annual amount to the company before advertising |
| 10 | Aft\_sales | Annual amount to the company after advertising |

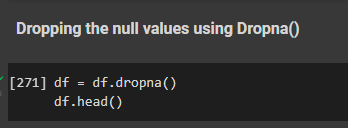
Page | 4

**3.2 Data Cleaning:**

Data cleaning includes removing all the null values from the dataset so that model runs more accurately.



We can remove the null values by either dropping those particular rows,



Or null or nan values can also be

->replaced by max value/ mean/ mode || median

*# Replace missing values with maximum value*

print(ldf.LoanAmount.head(40))

mdf3=ldf.LoanAmount.fillna(ldf.LoanAmount.min())

print(mdf3.head(40))

Page | 5

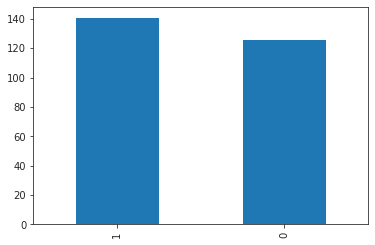
### **3.3 Data Augmentation**

Data augmentation is a set of techniques to artificially increase the amount of data by generating new data points from existing data. This includes making small changes to data or using deep learning models to generate new data points. In this dataset we add new every month, its like updating of dataset. To perform this task, we use data augmentation. By adding data we can get perfect prediction.

### **3.4 Data Visualization**

**Bar plot:**

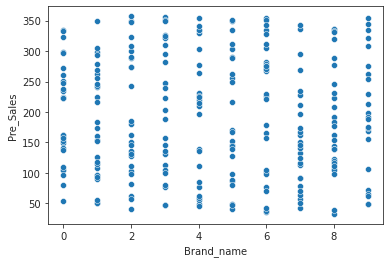
df.location.value\_counts().plot.bar()



It, represents the location of the advertising or marketing either it is Rural or Urban

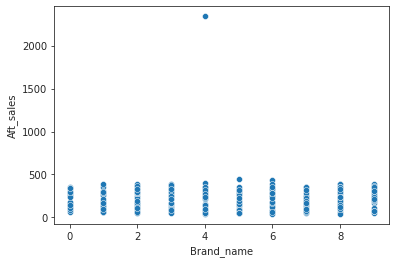
**Scatterplot:**

sns.scatterplot(x=df.Brand\_name,y = df1.Pre\_Salescatterplot



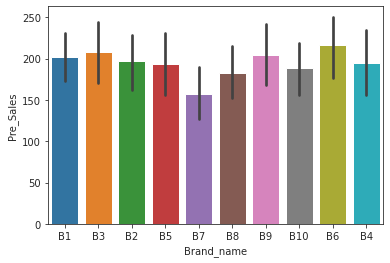
Page | 6

sns.scatterplot(x=df.Brand\_name,y = df1.Aft\_sales)



This graph helps us to know the sales of a particular brand when it is advertised.

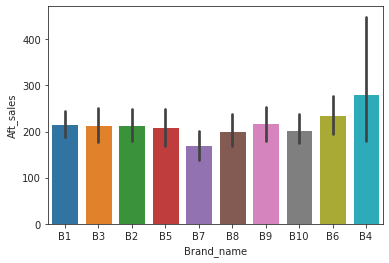
sns.barplot(x='Brand\_name',y='Pre\_Sales',data=df1);



It shows the sales of particular brand and it's sales before marketing.

Page | 7

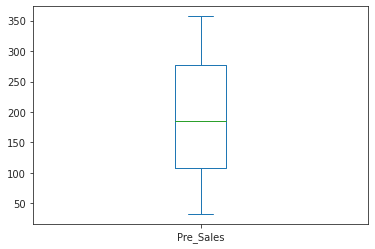
sns.barplot(x='Brand\_name',y='Aft\_sales',data=df1);



It shows the sales of particular brand and it's sales after marketing.

Box plot:

df1.Pre\_Sales.plot.box()

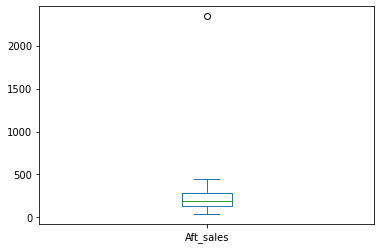


Page | 8

**OUTLIER HANDLING:**

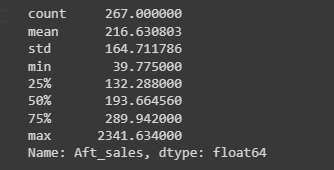
**Outliers are the points that differ from the remaining datapoints with very high or very low values.**

df.Aft\_sales.plot.box()



as we observed, here in the graph we can see e OUTLIER.

df['Aft\_sales'].describe()



Here we got the max and min values of datapoints of these particular sales.

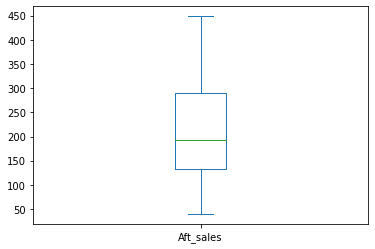
Now as observed in the result the mean is around 230 but max value is given as 2300,

So we will replace them.

Page | 9

df.loc[df['Aft\_sales']>=500,'Aft\_sales']=230

df.Aft\_sales.plot.box()



Now we can see that the outlier is removed, cause we change the values which lie in higher range(>500) to a constant 230

Page | 10

import matplotlib.pyplot as plt

from matplotlib.ticker import FormatStrFormatter

ax=sns.scatterplot(x='Pre\_Sales',y = 'Aft\_sales',hue="Brand\_name",data = df,s=80)

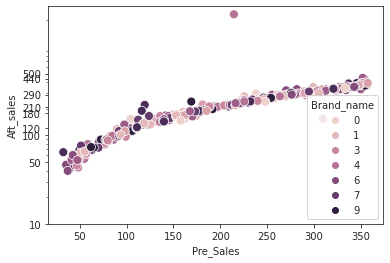
ax.set\_yscale('log')

ax.set\_yticks([10,50,100,120,180,210,290,440,500])

# ax.set\_yticklabels([0.03, 0.1, 0.3, 1, 3, 10, 30, 100, 300, 1000])

ax.yaxis.set\_major\_formatter(FormatStrFormatter('%g'))

plt.show()



here, the comparison between previous sales and after sales of all the brand names.

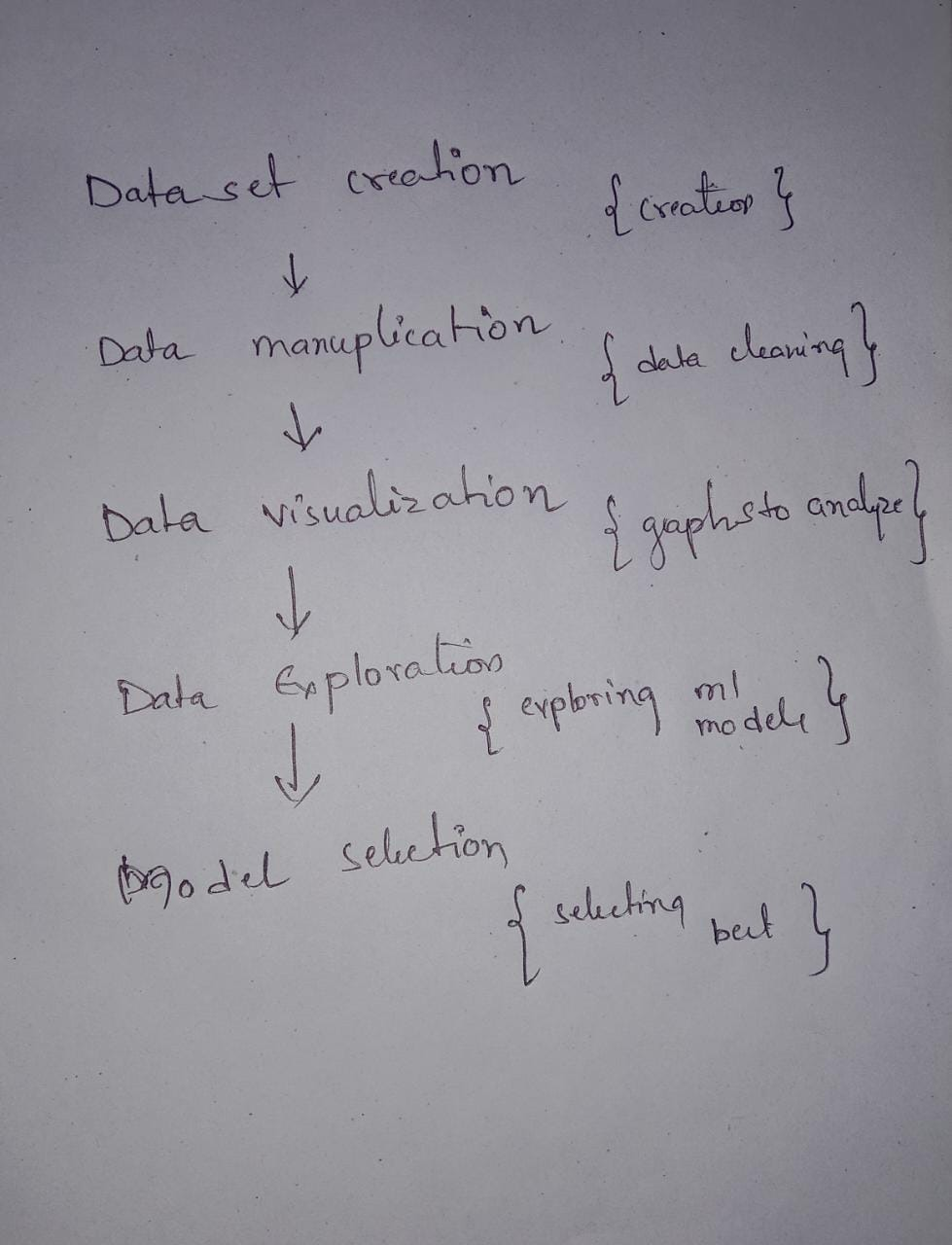
Page | 11

**4. METHODOLOGY**

### **4.1 Procedure to solve the given problem**:

### The procedure to solve the given problem statement is to train the model with appropriate ML models and out of those we need to select the best model to run the model and fit it.

### **4.2 Model Architecture :**



Page | 12

**Code:**

import pandas as pd

df=pd.read\_csv('/content/dafeml.csv')

df1=pd.read\_csv('/content/dafeml.csv')

df.head()

#null value -columwise

df.isna().sum()

df.columns

Index(['Brand\_name', 'Year', 'TV', 'Radio', 'Social Media', 'Influencer','Sponsering', 'location', 'Pre\_Sales', 'Aft\_sales'], dtype='object')

Finding the null values in the dataset using isnull() function in Heatmap function

import matplotlib.pyplot as plt

import seaborn as sns

def get\_heatmap(df):

    #This function gives heatmap of all NaN values

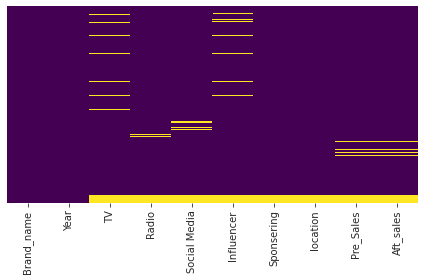
    plt.figure(figsize=(6,4))

    sns.heatmap(df.isnull(), yticklabels=False, cbar=False, cmap='viridis')

    plt.tight\_layout()

    return plt.show()

get\_heatmap(df)



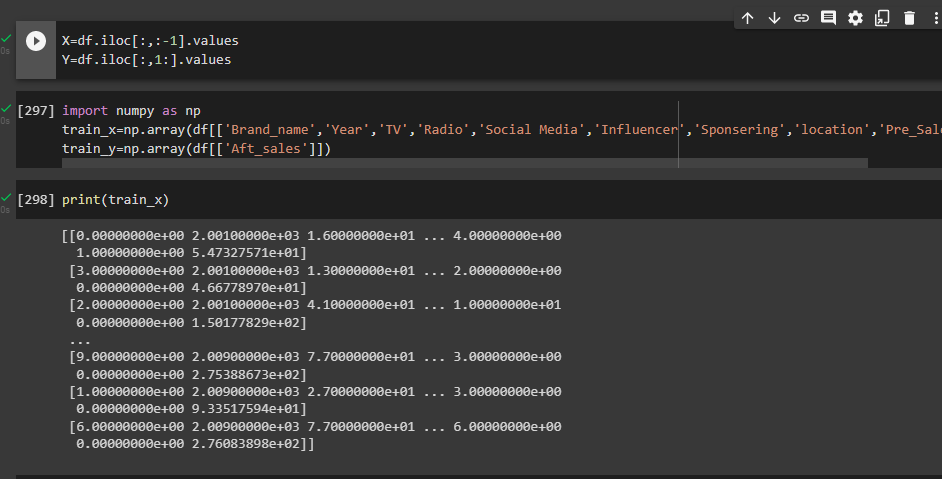
Dropping the null values using Dropna()

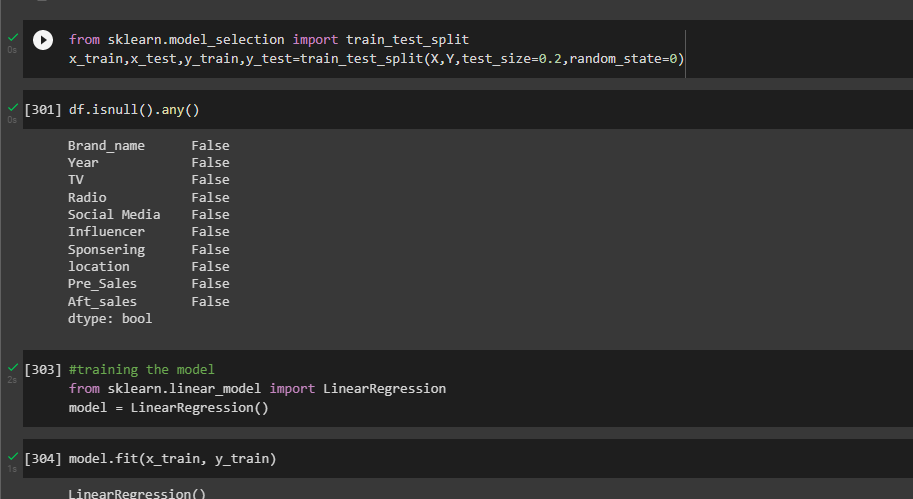
df = df.dropna()

df.head()

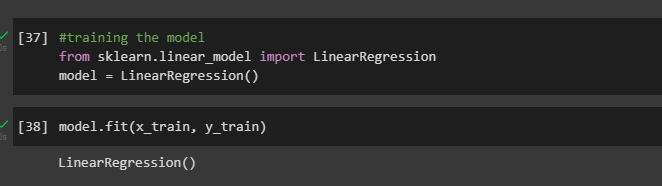
Page | 13

**Splitting Data after removing outliers into train and test data:**



****

**SELECTING APPROPRIATE MODEL:**

****

Page | 14

**4.3 Software description :**

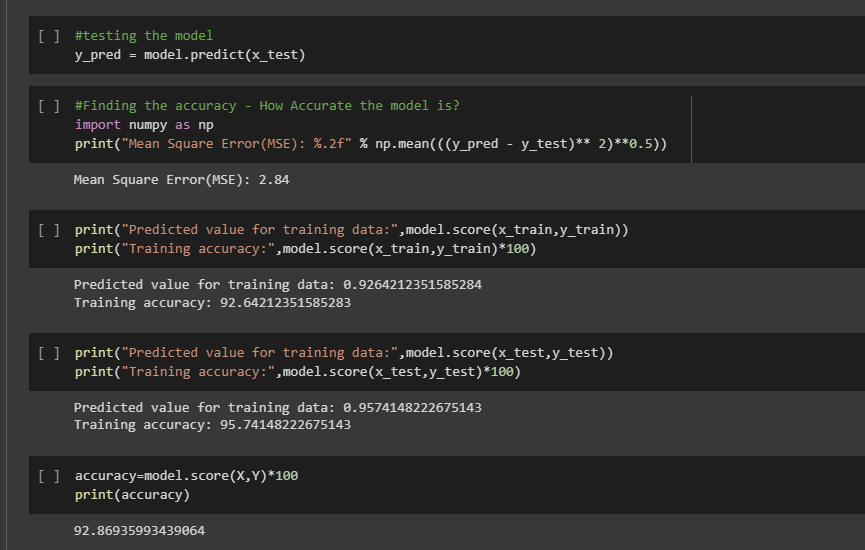
In our program we used python 3 programming language in Jupyter Notebooks for data visualization and model fitting. Python is an interpreted, object-oriented, high-level programming language with dynamic semantics. Its high-level built in data structures, combined with dynamic typing and dynamic binding, make it very attractive for Rapid Application Development, as well as for use as a scripting or glue language to connect existing components together. Python supports modules and packages, which encourages program modularity and code reuse. We installed certain packages like pandas, numpy for data visualization and matplotlib for generating graphs. The Python interpreter and the extensive standard library are available in source or binary form without charge for all major platforms, and can be freely distributed.

Page | 15

**5. RESULTS AND DISCUSSION**

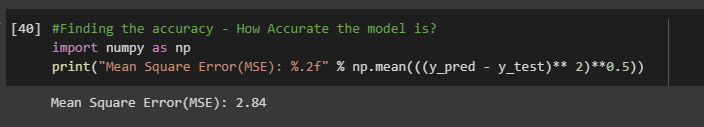
This is a regression-based model which will predict the increase in sales of brand after advertising their products.

**Accuracy:**

****

For Linear Regression, we use the **Mean Squared Error (MSE)** cost function, which is the average of squared error occurred between the predicted values and actual values. It can be written as:

Linear Regression in Machine Learning

****

Page | 16

**correlation matrix:**

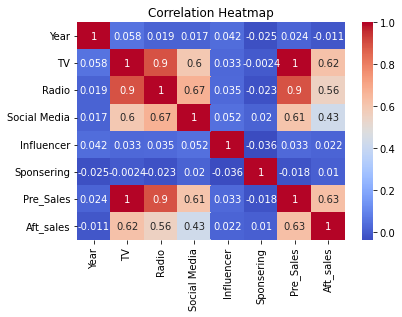
A correlation matrix is simply a table which displays the [correlation](https://corporatefinanceinstitute.com/resources/knowledge/finance/correlation/) coefficients for different variables. The matrix depicts the correlation between all the possible pairs of values in a table. It is a powerful tool to summarize a large dataset and to identify and visualize patterns in the given data.

A correlation matrix consists of rows and columns that show the variables. Each cell in a table contains the correlation coefficient.

sns.heatmap(df1.corr(), annot=True, cmap="coolwarm")

plt.title("Correlation Heatmap")

plt.show()

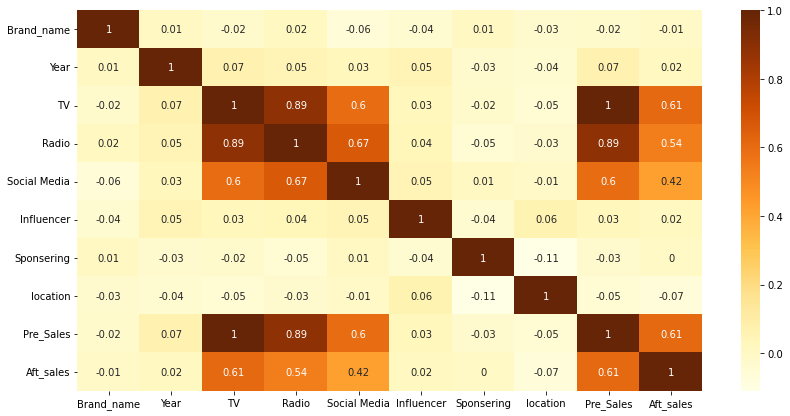


Page | 17

correlation = df.corr().round(2)

plt.figure(figsize = (14,7))

sns.heatmap(correlation, annot = True, cmap = 'YlOrBr')



This shows that,

dark orange ->will have high correlated

light orange ->2nd highest

Pale orange - > least correlated

Page | 18

# **6. CONCLUSION AND FUTURE SCOPE**

The conclusion of our analysis is that, advertising a key role in increasing of sales of any product.

From the dataset we have took very few categories of advertising platforms but there are many other platforms also.

As seen in the visualization product sales before marketing and after marketing are changed after the model being trained considering the amount spent on marketing.

The future scope for the project will be that as the main aim of this project is to know the relation among the sales ang the mode of marketing platform, the users can analyze and choose best platform so that they get maximum profits.

By using this they need not invest their money in all the modes of marketing but instead choose a best one.

In other mays it will also help the user to know how much average amount need to be spent for marketing.

Page | 19

# **REFERENCES**

While preparing the dataset, as we need few attributes like sales of companies in different years we preferred to

[Dummy Marketing and Sales Data | Kaggle](https://www.kaggle.com/datasets/caf07f0f20b7756940b5eccc22de4ec4e95cb1b4fe7d328d491db55207e66df1)

The following dataset.

As we used Linear Regression, to know more in detail about it, we had a look at

[Linear Regression in Machine learning - Javatpoint](https://www.javatpoint.com/linear-regression-in-machine-learning)

Page | 20