Savitribai Phule Pune University, PUNE

**A**

**PROJECT REPORT**

**ON**

**“T-Series DVD Sell Analysis”**

**Submitted by:**

**TEAM NO.: 21**

**Mr. Malwale Sharad Prabhakar Roll No.: COTB43**

**Mr. Patil Shivam Dinesh Roll No.: COTB44**

**Mr. Padul Umesh Chandrakant Roll No.: COTB45**



**SINHGAD ACEDAMY OF ENGINERRING, KONDHWA**

**DEPARTMENT OF COMPUTER ENGINERRING**

**Academic Year 2020-21**

**ABSTRACT**

**T-Series DVD Sell Analysis** uses the Data Analysis for operation. By using T-Series DVD Sell Analysis it is possible to analyze – Large and Complex data like T-series DVD Sell records.

More complex and hard to handle data can be analyzed and classified in easily viewable format.

We can represent data more effectively and in proper format for handling such a large data.

More emphasis can be done on data throughout the project operation.

**INDEX**

|  |  |  |
| --- | --- | --- |
| **Sr. No.** | **Topic Name** | **Page No.** |
| **1** | **Introduction**   * 1. Overall Description   2. Scope of the Project   3. Limitation of the Project   4. Feasibility Study | **1**  **1**  **2**  **2** |
| **2** | **Requirement Analysis**  2.1 Programming Language  2.2 Environment | **3**  **3** |
| **3** | **Project Planning**  3.1 Project Plan  3.2 Risk Assessment | **4**  **4** |
| **4** | **Modeling**  4.1 Existing System  4.2 Drawbacks of Existing System | **5**  **5** |
| **5** | **Coding** | **6** |
| **6** | **Screenshots** | **7** |
| **7** | **Figure** | **12** |
| **8** | **Applications and Future Scope** | **17** |
| **9** | **Conclusion** | **18** |
| **10** | **References** | **19** |
|  |  |  |

**FIGURE INDEX**

|  |  |  |
| --- | --- | --- |
| **Fig. No.** | **Figure Name** | **Page No.** |
| 1 | Screenshot (6.1) Sold in 2016 |  |
| 2 | Screenshot (6.2) Sold in 2017 |  |
| 3 | Screenshot (6.3) Sold in 2018 |  |
| 4 | Screenshot (6.4) Sold in 2019 |  |
| 5 | Screenshot (6.5) Sold in 2020 |  |
| 6 | Graph (7.1) Output 1 |  |
| 7 | Graph (7.2) Output 2 |  |
| 8 | Graph (7.3) Output 3 |  |
| 9 | Graph (7.4) Output 4 |  |
| 10 | Graph (7.5) Output 5 |  |

1. **INTRODUCTION**

Engineering is used to find the solutions for the biggest problems. The large and complex data is being produced with every second of time in this world.

To provide a proper manner and efficient way to represent such large and complex data, we need to use programming languages for better representation.

By implementing the project, the developers and data scientists will be able to manipulate, manage and modify data more effectively.

For implementing this project, we used T-Series DVD sold records for representing DVDs sold daily, monthly and yearly. It provided overall view for large data of the last 4 years (Nov-2016 to Oct-2020).

* 1. **Overall Description**

The project includes features like:

1. Data Manipulation.
2. Data Management.
3. Data Modification.
4. Data Analyzation.
5. Data Representation.
   1. **Scope of the Project**

Scope for this project is to emphasis on the data and its appropriate and effective representation. The data manipulation with the help of importing through the files format like excel, csv, etc. The data management with the help of arrangement of the data like Data Frames. The data modification with the help of libraries of the programming languages like Pandas from Python. The data analyzation with the help of algorithms like built-in Python Algorithms used for Data Science. The data representation with the help of graphs like heatmaps (Matplotlib Library) provided by Python.

* 1. **Limitation of the Project:**

This project won’t be appropriate and efficient for Big Data. This representation of data requires raw data in appropriate format.

* 1. **Feasibility Study:**

To determine the usability and feasibility of the proposed project, We further study:

1. **Technical Feasibility:**

The Project can represent the data effectively irrespectively of its complex structure.

1. **Operational Feasibility:**

Can be operated easily through regular practice.

1. **Economical Feasibility:**

The throughput is not only good enough but also additionally time and money is saved.

1. **REQUIREMENT ANALYSIS**
   1. **Programming Language:**

A language which support operations related to data handling and manipulation is suitable for the data analysis project.

Languages like R was main choice of the data scientist for the data operation processes. Now, Python has taken the place of R language with additional features and ease of use.

Python provides vast variety of libraries for implementation of such projects which has emphasis on data and its analyzation.

* 1. **Environment:**
     1. Hardware Requirements:

1. Processor : Intel Dual Core
2. RAM : 2 GB
3. Hard Disk : 50 GB
4. Monitor : VGA/SVGA/HDMI
5. Keyboard : 100 Keys
6. Mouse : 2 buttons/ 3 buttons
7. Graphics Card : 2 GB
   * 1. Software Requirement:

As per the requirements for data emphasized project we required the R Studio for implementation of graphs and Python Language is used for plotting and manipulating Graphs and Data respectively.

1. Operating System : Windows 8/10 Professional Edition
2. Front End : Python 3
3. Back End : Python 3.8
4. Software Technology : R Studio
5. Data File Format : .csv/ .xlsx (Microsoft Excel)
6. **PROJECT PLANNING**
   1. **Project Plan:**

|  |  |  |  |
| --- | --- | --- | --- |
| **Sr. No.** | **Module** | **Start Date** | **End Date** |
| 1 | Searching for project topics. | 11/11/2020 | 13/11/2020 |
| 2 | Identifying needs and benefits. | 13/11/2020 | 15/11/2020 |
| 3 | Discussion with project guide. | 18/11/2020 | 18/11/2020 |
| 4 | Collecting information about project and identify project constraints. | 18/11/2020 | 21/11/2020 |
| 5 | Requirement and system analysis. | 21/11/2020 | 24/11/2020 |
| 6 | Data files preparation. | 25/11/2020 | 26/11/2020 |
| 7 | Getting started with Python. | 26/11/2020 | 28/11/2020 |

|  |  |  |
| --- | --- | --- |
| **Date** | **Project Module** | **Project Milestone** |
| 25/11/2020 to 26/11/2020 | Data Files Creation | Data Files Created |
| 26/11/2020 to 27/11/2020 | Coding | Code Completed |
| 27/11/2020 to 28/11/2020 | Plotting Graphs | Graphs Plotted |
| 28/11/2020 to 29/11/2020 | Documentation | Documentation Completed |

* 1. **Risk Assessment:**

Every challenge comes with the risks which may lead severe damage to the project or environment/ system.

Some risks may effect efficiency only but those type of risks also should be considered and reducing their respective damages is beneficial for the project.

1. **MODELING**
   1. **Existing System:**

The existing system can only store data in tabular format. It is just a record of some data.

* 1. **Drawbacks of Existing System:**
* Data is complex to look at.
* Data representation is not possible with ease.
* Not Suitable for large data.
* Isolation of data is harder.
* Manipulation of data takes more effort and time.
  1. **Need for the System:**

Existing system has lot of limitation as described above. To remove these limitations and adding some features we require the analysis of the data.

* Plotted graphs is easier to understand.
* Data can be overlooked easily.
* Flexibility of data representation is available.
* Large data is covered.

1. **CODING**

import pandas as pd

import seaborn as sns

import matplotlib.pyplot as plt

data = pd.read\_csv(r"C:\Users\8umes\Desktop\mm\_dd\_all\_ccssvv.csv")

data['Months16'] = pd.to\_datetime(data.Months16, format='%b', errors='coerce').dt.month

heatmap\_data1 = pd.pivot\_table(data, values='Values16', index=['Dates16'], columns='Months16')

plt.figure(figsize=(20, 10))

sns.heatmap(heatmap\_data1, cmap="coolwarm",annot=True, fmt = 'g')

plt.title("DVD Sold in Year 2016", fontsize =20)

plt.show()

data['Months17'] = pd.to\_datetime(data.Months17, format='%b', errors='coerce').dt.month

heatmap\_data2 = pd.pivot\_table(data, values='Values17', index=['Dates17'], columns='Months17')

plt.figure(figsize=(20, 10))

sns.heatmap(heatmap\_data2, cmap="coolwarm",annot=True, fmt = 'g')

plt.title("DVD Sold in Year 2017", fontsize =20)

plt.show()

data['Months18'] = pd.to\_datetime(data.Months18, format='%b', errors='coerce').dt.month

heatmap\_data3 = pd.pivot\_table(data, values='Values18', index=['Dates18'], columns='Months18')

plt.figure(figsize=(20, 10))

sns.heatmap(heatmap\_data3, cmap="coolwarm",annot=True, fmt = 'g')

plt.title("DVD Sold in Year 2018", fontsize =20)

plt.show()

data['Months19'] = pd.to\_datetime(data.Months19, format='%b', errors='coerce').dt.month

heatmap\_data4 = pd.pivot\_table(data, values='Values19', index=['Dates19'], columns='Months19')

plt.figure(figsize=(20, 10))

sns.heatmap(heatmap\_data4, cmap="coolwarm",annot=True, fmt = 'g')

plt.title("DVD Sold in Year 2019", fontsize =20)

plt.show()

data['Months20'] = pd.to\_datetime(data.Months20, format='%b', errors='coerce').dt.month

heatmap\_data5 = pd.pivot\_table(data, values='Values20', index=['Dates20'], columns='Months20')

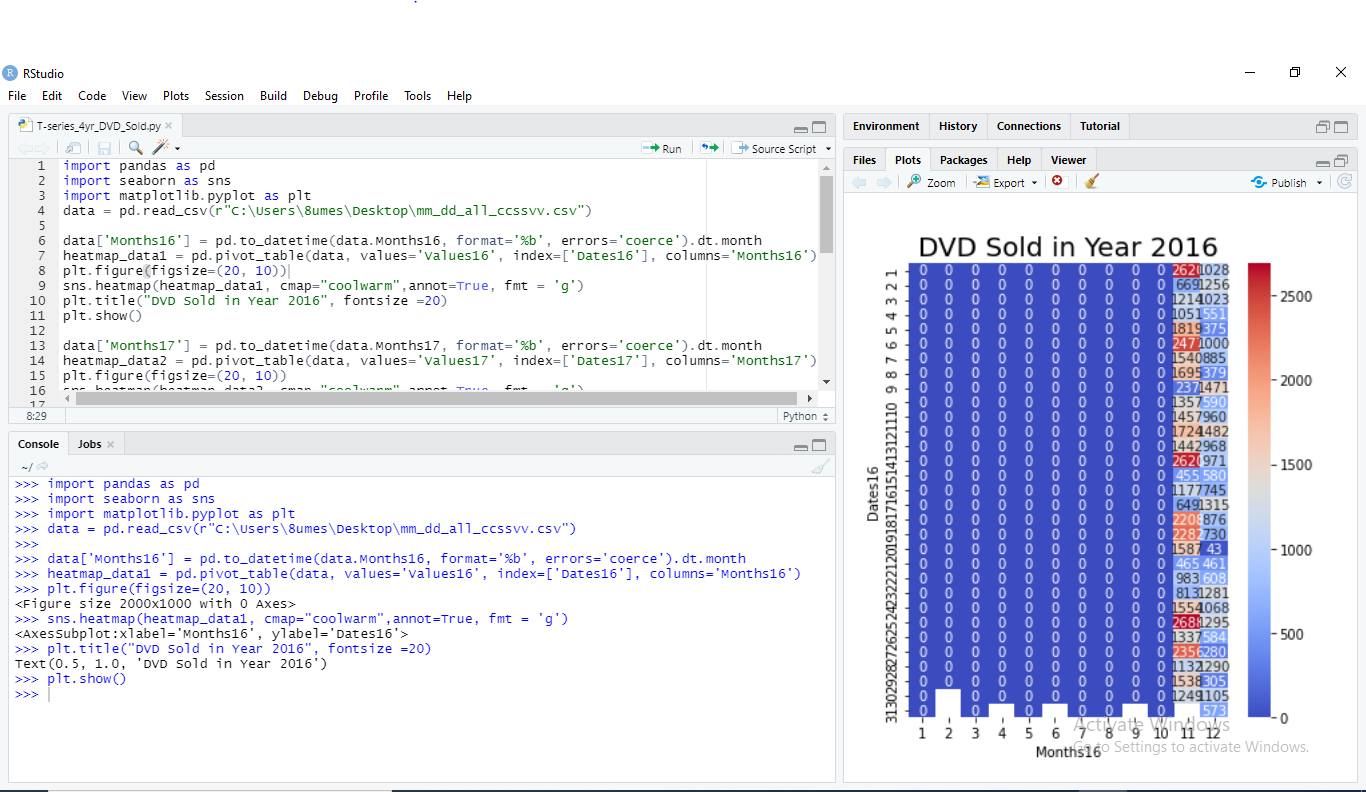
plt.figure(figsize=(20, 10))

sns.heatmap(heatmap\_data5, cmap="coolwarm",annot=True, fmt = 'g')

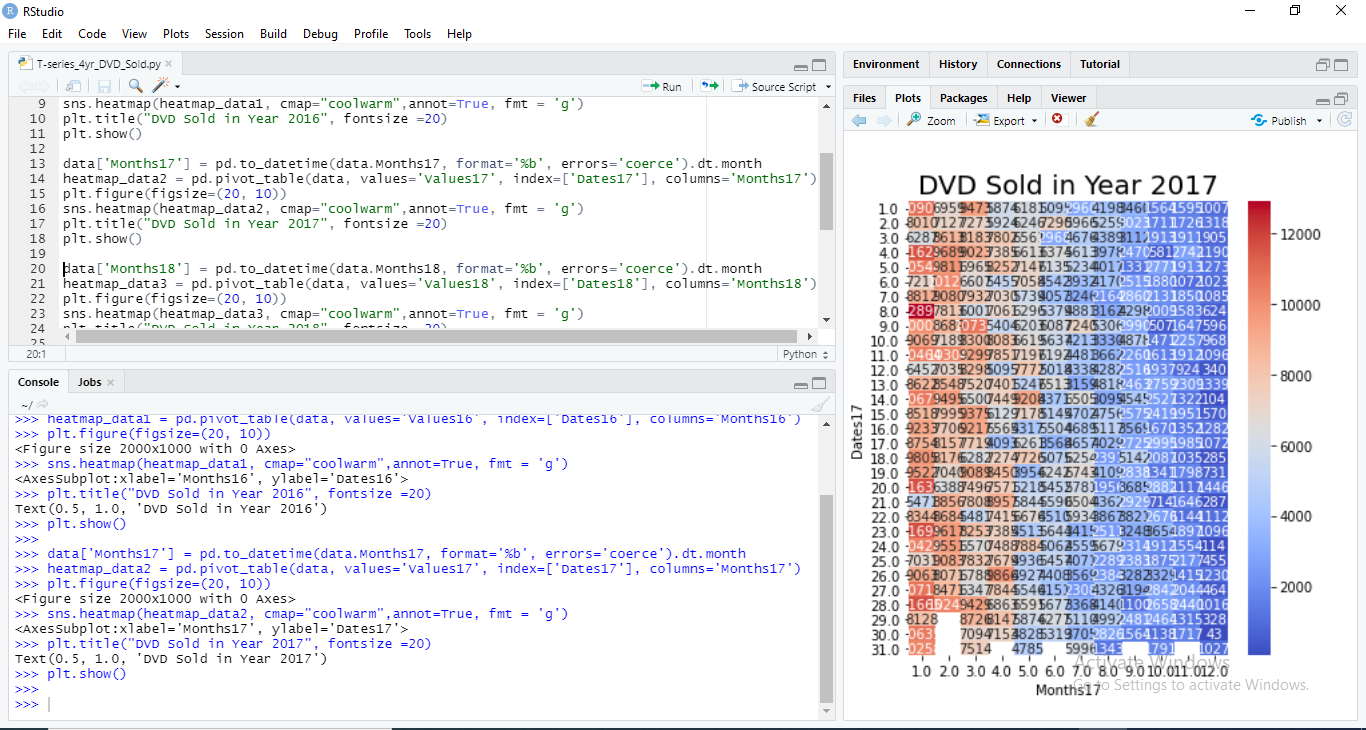
plt.title("DVD Sold in Year 2020", fontsize =20)

plt.show()

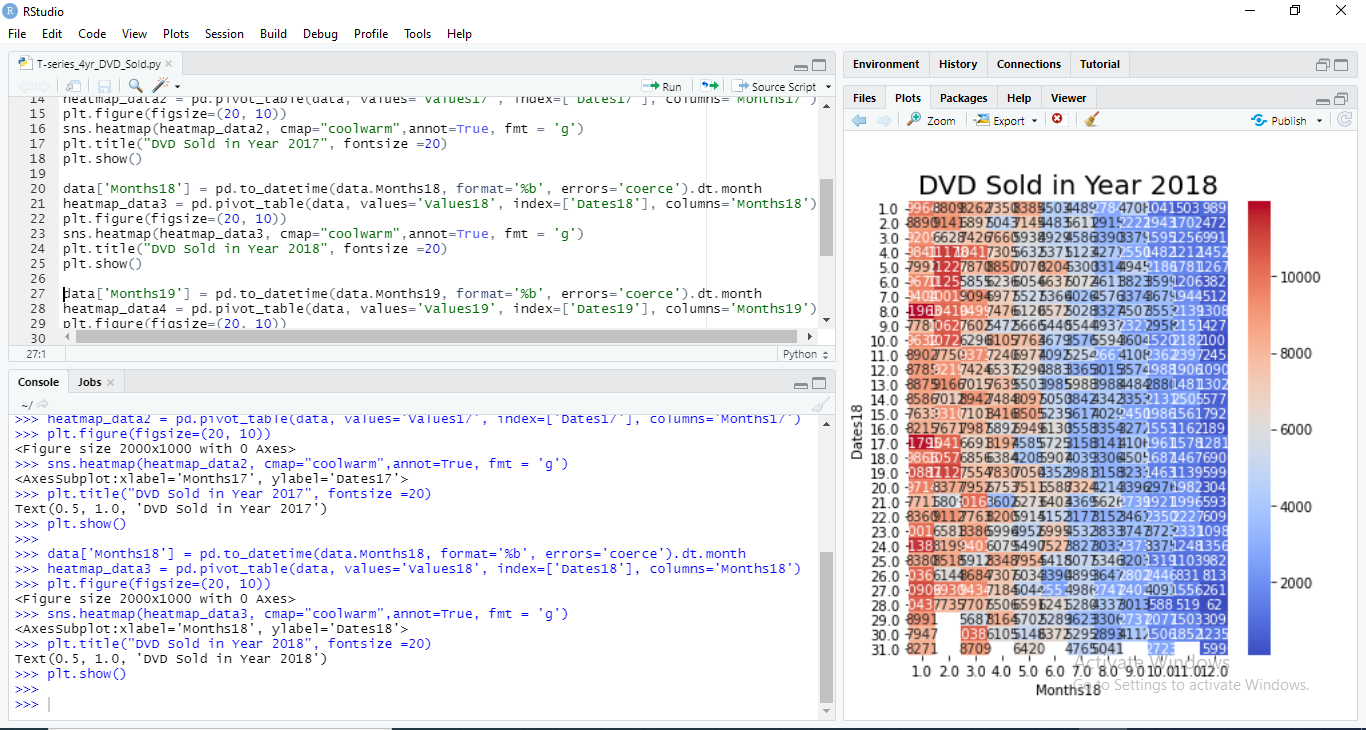
1. **SCREENSHOTS**



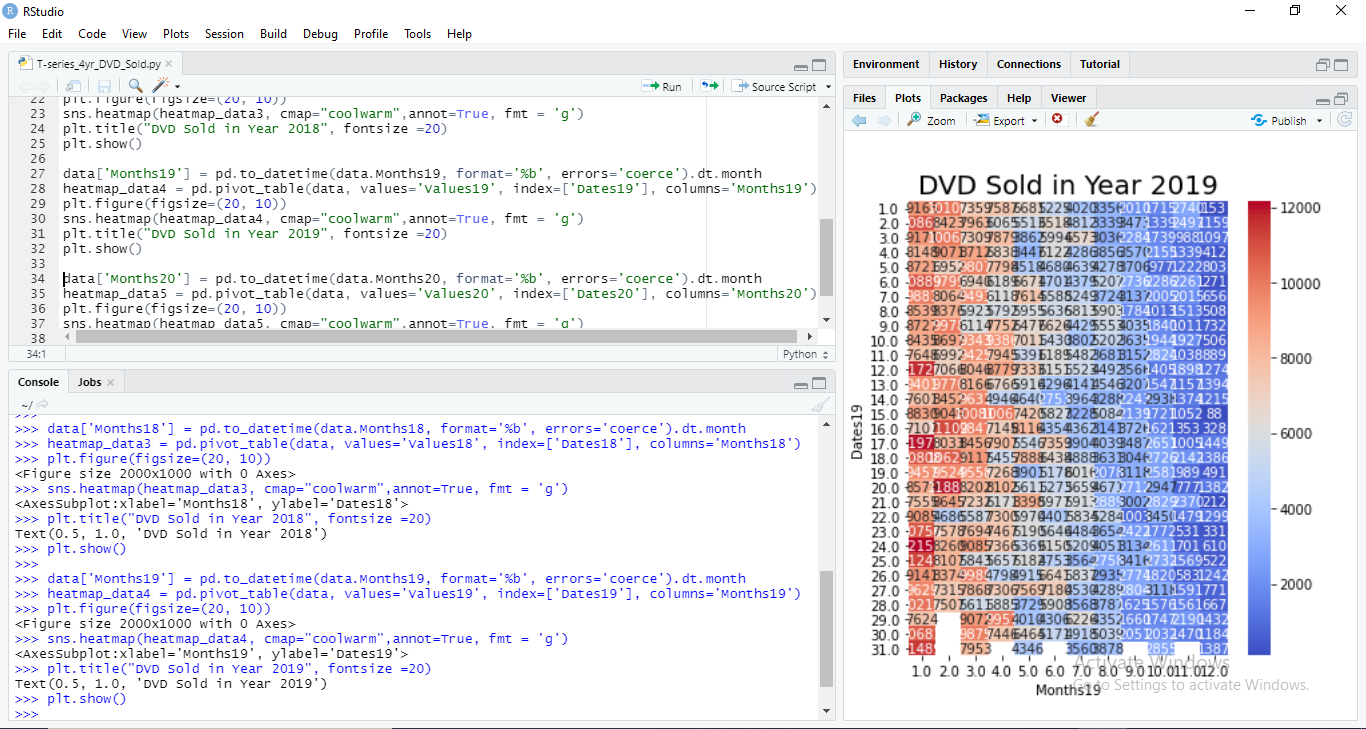
**Screenshot (6.1) Sold in 2016**



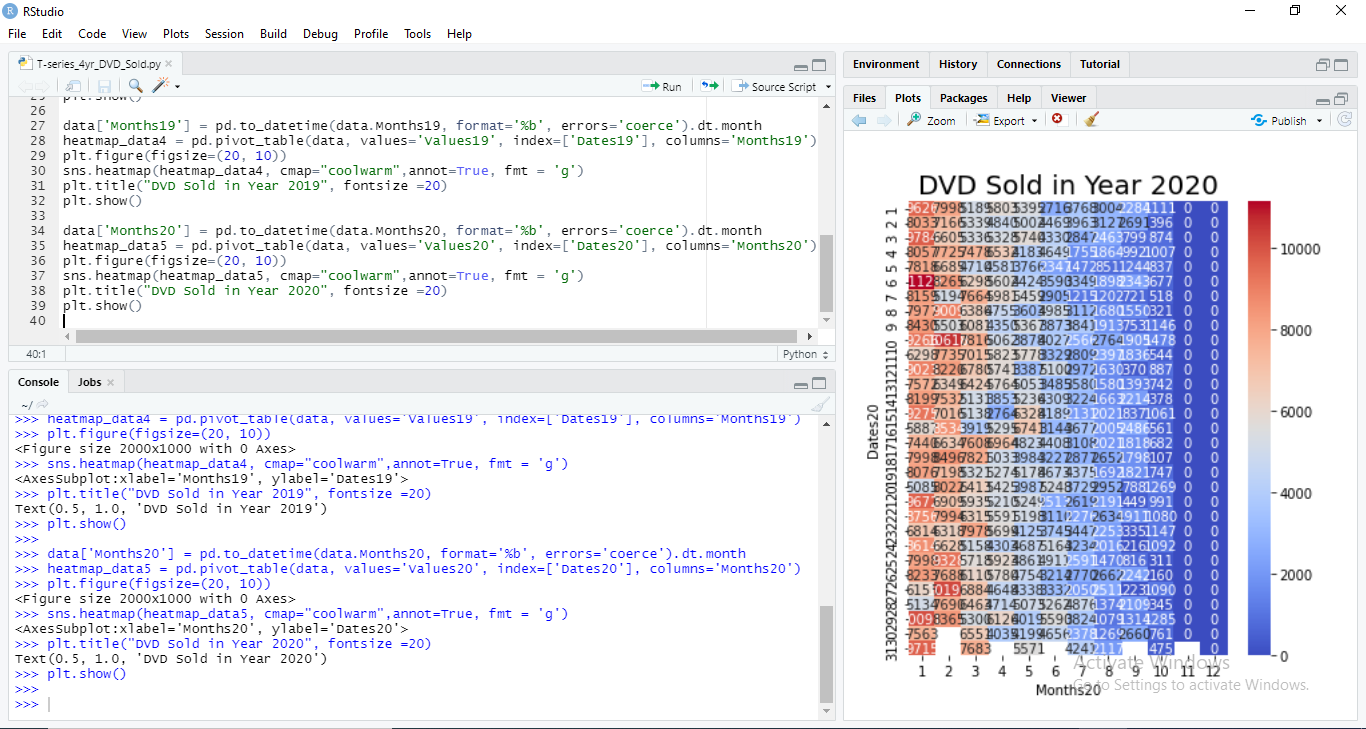
**Screenshot (6.2) Sold in 2017**



**Screenshot (6.3) Sold in 2018**

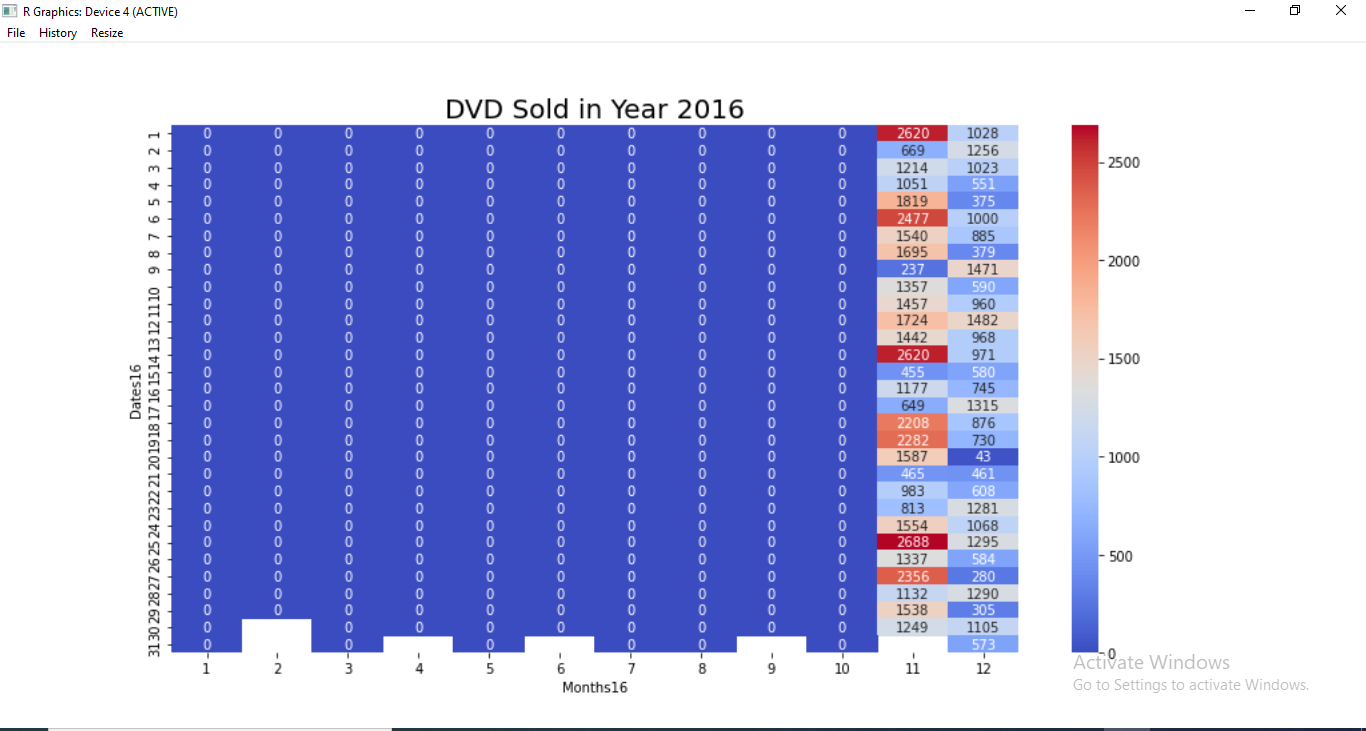


**Screenshot (6.4) Sold in 2019**

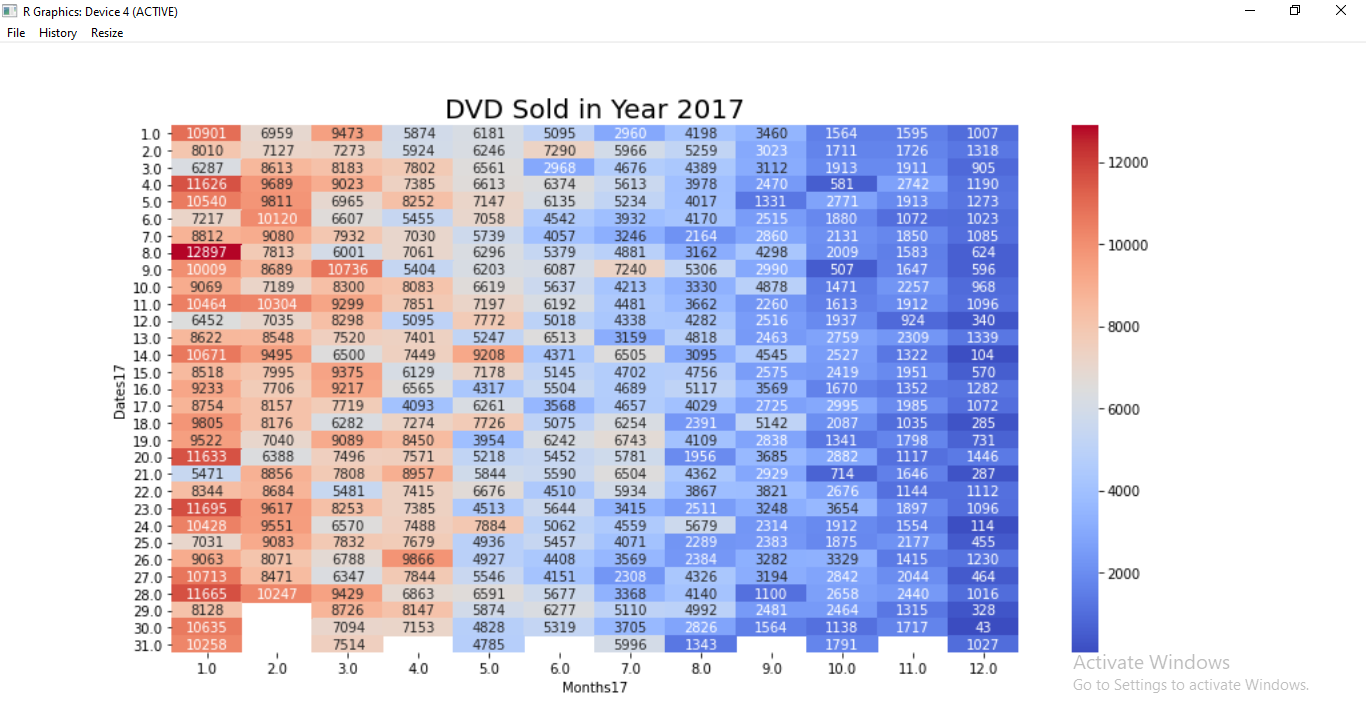


**Screenshot (6.5) Sold in 2020**

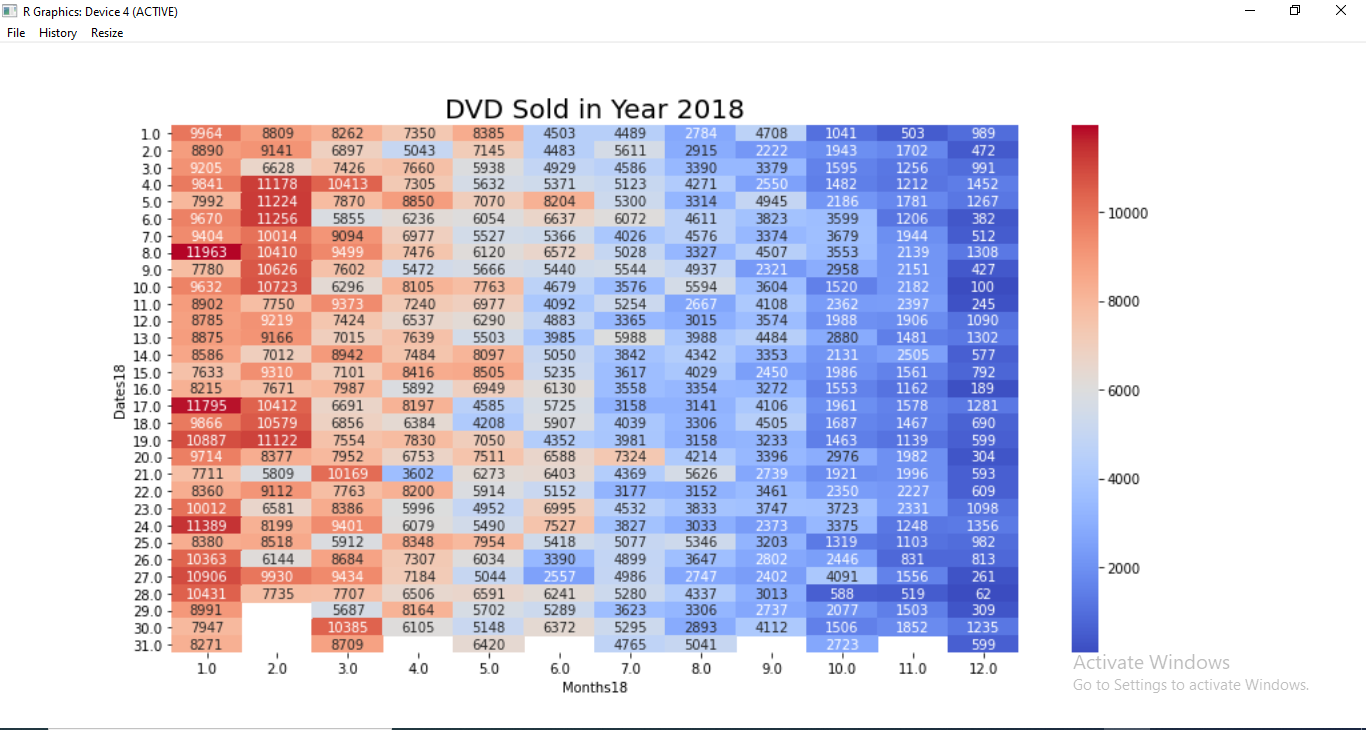
1. **FIGURE**



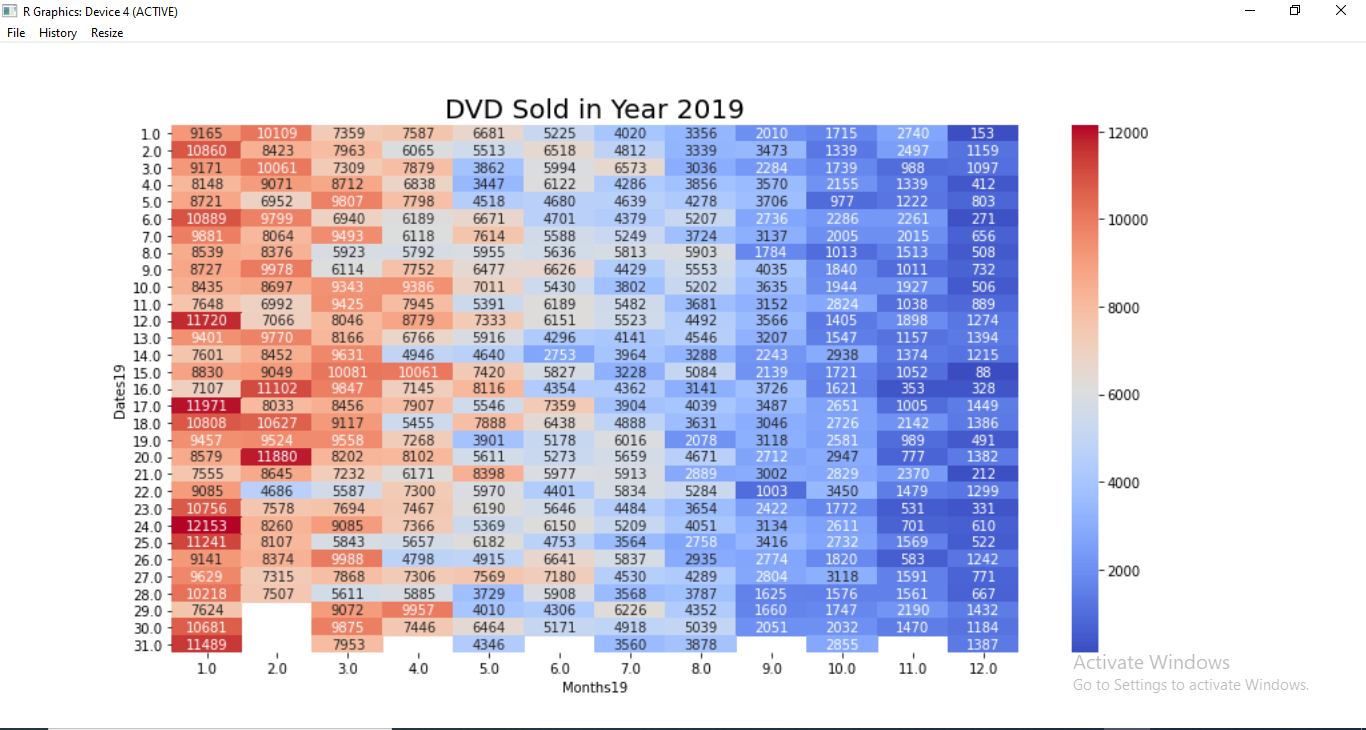
**Graph (7.1) Output 1**



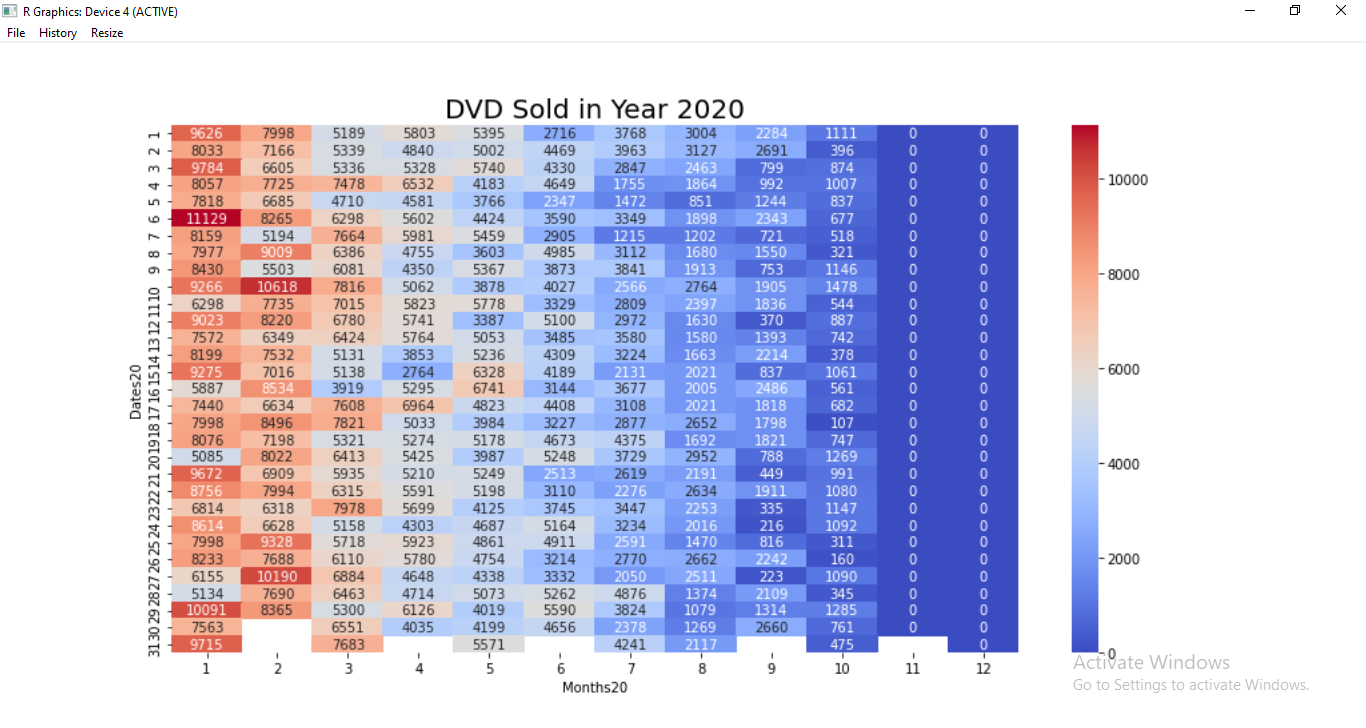
**Graph (7.2) Output 2**



**Graph (7.3) Output 3**



**Graph (7.4) Output 4**



**Graph (7.5) Output 5**

1. **APPLICATION AND FUTURE SCOPE**

Applications of the project are more focused on data representation and data manipulation. It also consists of conversion of raw data into structured and modifiable data. It can be further developed for future use like Big Data handling, tool for Data Scientist and for respective graphics and plotting operations performed on data for statistical use.

It may be developed for statistical applications like data mining for major operations. It may play bigger role in Information Technology.

1. **CONCLUSION**

The utilization of data libraries provided by Python helped in plotting graphs for raw data with environment as R Studio. Soft skills are developed for programming languages like Python and R with good experience.

This project is very useful plotting and representing the large data in single view ease.

Emphasis was on data and practical example of that was plotting which is implemented on T-Series DVD sold record.

1. **REFERENCES**

* <https://www.python.org/>
* <https://rstudio.com/>
* <https://www.r-project.org/>
* <https://pandas.pydata.org/>
* <https://matplotlib.org/>
* <https://seaborn.pydata.org/>