K L E F
KONERU LAKSHMAIAH EDUCATION FOUNDATION
(Deemed to be university estd, u/s, 3 of the UGC Act, 1956)
(NAAC Accredited "A" Grade University)

PROJECT REPORT

On

ANALYSIS AND VISUALIZATION OF US FERTILITY TRENDS

Submitted in partial fulfilment of the
Requirements for the award of the Degree of
Bachelor of Technology
In
Computer Science and Engineering
Under the esteemed guidance of

Dr. DEBRUP BANERJEE

By

TEJASWI REDDY. K

(DST-FIST Sponsored Department)

K L EDUCATION FOUNDATION

Green Fields, Vaddeswaram, Guntur District-522 502

2019-2020

K L EDUCATION FOUNDATION

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

(DST-FIST Sponsored Department)

CERTIFICATE

This is to certify that this project based lab report entitled **“ANALYSIS AND VISUALIZATION OF US FERTILITY TRENDS”** is a bonafide work done by **TEJASWI REDDY. K** in the course **18cs3262S DATA MODELING & VISUALIZATION** in partial fulfilment of the requirements for the award of Degree in Bachelor of Technology in **COMPUTER SCIENCEAND ENGNEERING** during the Even Semester of Academic year 2020-2021.

Dr. DEBRUP BANERJEE

Faculty in Charge

Dr. HARI KIRAN. V

Head of the Department

K L EDUCATION FOUNDATION

DEPT OF COMPUTER SCINCE AND ENGINEERING

(DST-FIST Sponsored Department)

DECLARATION

We hereby declare that this project based lab report entitled **"ANALYSIS AND VISUALIZATION OF US FERTILITY TRENDS"** has been prepared by us in the course **18CS3262S DATA MODELING & VISUALIZATION** in partial fulfilment of the requirement for the award of degree bachelor of technology in **COMPUTER SCIENCE AND ENGINEERING** during the Even Semester of the academic year 2020-2021. We also declare that this project-based lab report is of our own effort and it has not been submitted to any other university for the award of any degree.

Date: 04.11.2020

Place: Vijayawada

ACKNOWLEDGEMENT

Our sincere thanks to Dr. DEBRUP BANERJEE in the Lab for her outstanding support throughout the project for the successful completion of the work.

We express our gratitude to **Dr. P. Vidyullatha** Course Coordinator for **18CS3262S DATA MODELING & VISUALIZATION** course in the Computer Science and Engineering Department for providing us with adequate planning and support and means by which we can complete this project-based Lab.

We express our gratitude to **Dr. V. HARIKIRAN,** Head of the Department for computer science and Engineering for providing us with adequate facilities, ways and means by which we can complete this project-based Lab.

We would like to place on record the deep sense of gratitude to the honorable Vice Chancellor, K L University for providing the necessary facilities to carry the project-based Lab.

Last but not the least, we thank all Teaching and Non-Teaching Staff of our department and especially our classmates and our friends for their support in the completion of our project-based Lab.

Name of the student

TEJASWI REDDY. K

180030537

TABLE OF CONTENTS

CHAPTERS **PAGE NO**

ABSTRACT

The project **"ANALYSIS AND VISUALIZATION OF US FERTILITY TRENDS"** is based on the births taking place in United States from the years 1994 to 2003. It is being analyzed based on the births taking place in a particular month or day of the week or date of a month in a particular year. Here we are using the dataset "US_births_1994-2003_CDC_NCHS" to analyze the births taking place between the years 1994 and 2003(inclusive) to present a report on various analyses we have on the births taking place. This real-time dataset helps us in visualization of births taking place in the given range of time period.

INTRODUCTION

INTRODUCTION:

This report presents the data on U.S. births from the year 1994 to the year 2003 according to a wide variety of characteristics. Data is presented for characteristics including birth and fertility rate with date, day, month and year. Trends in fertility patterns and dates are described and interpreted. Descriptive tabulations of data reported on the birth certificates of the 4.5 billion births that occurred from 1994 to 2003 are presented.

PROBLEM DEFINITION:

This Project is about visualizing the Trends of the Birth Fertility Rate in the United States of America. It is used to demonstrate and visualize the number of births taking place in United States of America using Tableau and Python. Here we represented the number of births from various aspects.

There are some issues taking place at the time of delivery, like miscarriage and unhealthiness of the mother or the baby due to weakness and other reasons. There are even chances of the death of the baby or mother at the time of delivery. Such things may even lead to increase or decrease of birth rate which may become a problem in increment or decrement of population rate of the country.

SCOPE:

The U.S. birthrate fell again in 1997, to 3,788,235 births — representing a 2% drop from 1996. It's the lowest number of births, according to the visualized report. The numbers also sank the U.S. fertility rate to a record low. Not since 1986 has the U.S. seen so few babies born.

PURPOSE:

To demonstrate and visualize the births in United States of America using Tableau and Python. Here we represented the number of births from US between the years 1994 to 2003. Major number of births took place in the year 2003 of about more than 4million.

PROBLEM AND EXISTING TECHNOLOGY:

For Medical Hospitals:

1. The number of beds for pregnant ladies should me more in the months of August, September and July.

2. They can also have more staff on Tuesdays for conceiving babies.

3. They can staff more people during the middle of the month rather than on the end of the month for delivering babies.

For Baby Care and Product Industries:

The Baby care and safety products companies can adjust their marketing based on the months where more babies are being conceived. Here as we can see in the months of August, September and October more number of babies are delivered, so the companies can increase their marketing about baby products during these months.

For Contraceptive Industry:

In general Pregnancy lasts for about 280 days or 40 weeks that is around 9 months. So for instance the condom companies should try to increase their marketing 9-10 months before the months which have more number of births. Though we may require more data to deduce something like this like number of pregnancies, number of abortions, number of miscarriages and so on, these insights are based on the data we are currently using.

Here we can see that in the months of August, September and October the births are more. So the contraceptive companies should increase the marketing in the months of January, December and November.

PROPOSED SYSTEM:

The birth rate is one of the most basic and important measures in demography. But its relevance is not limited to just demographers. Birth rates affect public policy and budgeting for education and health systems, and can have major impacts on the well-being of a country's population. Population growth is based on four fundamental factors: birth rate, death rate, immigration, and emigration.

REQUIREMENTS AND ANALYSIS

PLATFORM REQUIRMENTS:

OPERATING SYSTEM: WINDOWS 7 or later version

TOOLS: PYTHON JUPITER NOTEBOOK,

TABLEAU

LANGUAGE: PYTHON

RAM: 1GB or more

HARD-DISK: 6GB or more

PROCESSOR: any INTEL processor

MODULE DESCRIPTION:

This module makes us understand the statistical birth fertility trends at the United States from the year 1994 to the year 2003. We have analyzed the given dataset and visualized it clearly to understand about the birth rate of US with the provided attributes like date, day, month and year.

DESIGN AND IMPLEMENTATION

ALGORITHMS:

IN JUPYTER:

Step 1: Open jupyter notebook.

Step 2: import the packages whatever we need for this project.

Step 3: read input data in csv

Step 4: Display the data present in the input file

Step 5: Now visualize the data representing the births from 1994 to 2003.

IN TABLEAU:

Step 1: Open Tableau Desktop

Step 2: Click on the text file or Microsoft excel based on the data

Step 3: Import dataset

Step 4: Perform the visualization based on problem statement and drag the values into rows and columns whatever we need.

Step 5: Apply different colors, shapes to the visualization for understanding easily.

Step 6: Rename the sheets.

Step 7: Create the Dashboard.

Step 8: Save the Book.

PSEUDO CODE:

```
import numpy as np

import pandas as pd

import seaborn as sns

import matplotlib as plt

from sklearn.preprocessing import LabelEncoder

from sklearn.impute import SimpleImputer

import mglearn

import plotly

# connected=True means it will download the latest version of plotly javascript library.

plotly.offline.init_notebook_mode(connected=True)

import plotly.graph_objs as go

import plotly.figure_factory as ff

import cufflinks as cf
import warnings

warnings.filterwarnings('ignore')

//importing csv file

birth1 = pd.read_csv(r'C:\Users\tejaswi\Downloads\dv\US_births_1994-2003_CDC_NCHS.csv')

birth1.describe()

birth1.info()

birth1.isnull().sum()

//bar chart for births in months

import plotly.express as px

#birth_month= px.birth.gapminder()

fig = px.bar(birth1, x='month', y='births')

fig.show()

//bar chart for births in years

import plotly.express as px

fig = px.bar(birth1, x='year', y='births')

fig.show()

//bar chart for births in days of a week
```

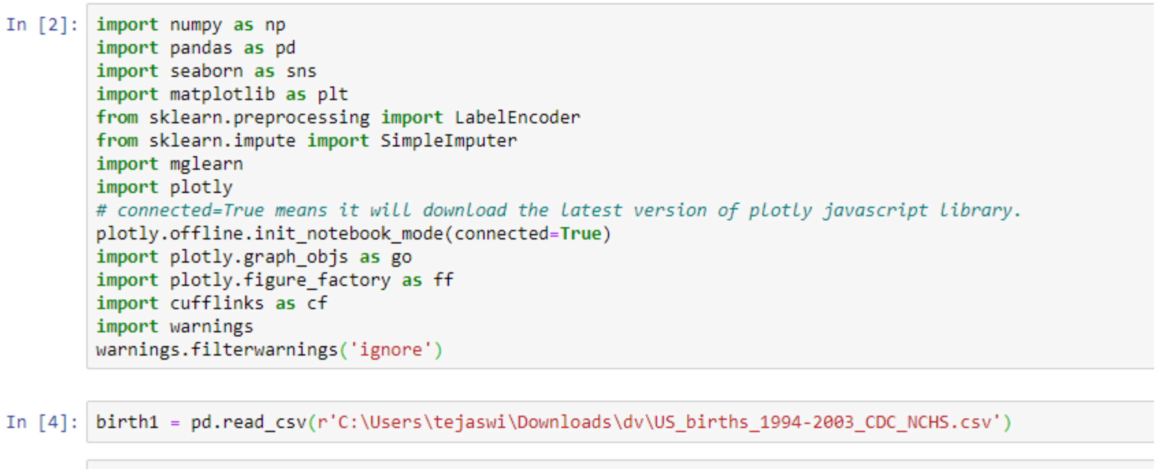
```
#birth_month= px.data.gapminder()

fig = px.bar(birth1, x='day_of_week', y='births')

fig.show()

//bar chart for births in date of month

#birth_month= px.data.gapminder()

fig = px.bar(birth1, x='date_of_month', y='births')

fig.show()

//pie chart to represent the births in month

import plotly.express as px

fig = px.pie(birth1, values='births', names='month')

fig.update_traces(textposition='inside')

fig.update_layout(uniformtext_minsize=12, uniformtext_mode='hide')

fig.show()

//pie chart to represent the births in days of week

import plotly.graph_objects as go

fig = go.Figure(data=[go.Pie(labels=birth1.day_of_week, values=birth1.births,
textinfo='label+percent', insidetextorientation='radial')])
```

```
fig.show()

//pie chart to represent the births in dates of month

import plotly.graph_objects as go

fig = go.Figure(data=[go.Pie(labels=birth1.date_of_month, values=birth1.births,
textinfo='label+percent', insidetextorientation='radial')])

fig.show()

//tree map to represent the births in month

import plotly.express as px

fig = px.treemap(birth1, path=['month'], values='births',

                 color='births', hover_data=['births'],

                 color_continuous_scale='RdBu',

                 color_continuous_midpoint=np.average(birth1['births'], weights=birth1['births']))

fig.show()

//tree map to represent the births in days of week

import plotly.express as px

import numpy as np

#df = px.data.gapminder().query("year == 2007")
```

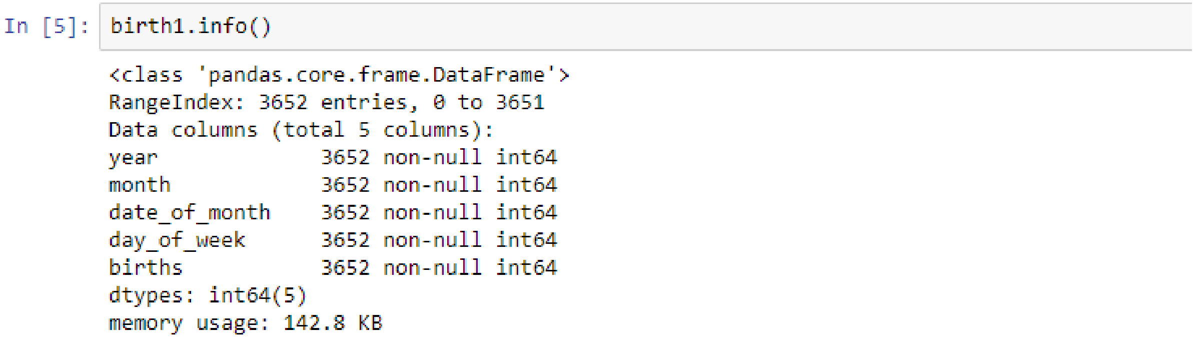
```
#df["world"] = "world" # in order to have a single root node

fig = px.treemap(birth1, path=['day_of_week'], values='births',

          color='births', hover_data=['births'],

          color_continuous_scale='RdBu',

          color_continuous_midpoint=np.average(birth1['births'], weights=birth1['births']))

fig.show()

//tree map represent the births in dates of month

import plotly.express as px

import numpy as np

#df = px.data.gapminder().query("year == 2007")

#df["world"] = "world" # in order to have a single root node

fig = px.treemap(birth1, path=['date_of_month'], values='births',

          color='births', hover_data=['births'],

          color_continuous_scale='RdBu',

          color_continuous_midpoint=np.average(birth1['births'], weights=birth1['births']))

fig.show()
```

SCREENSHOTS

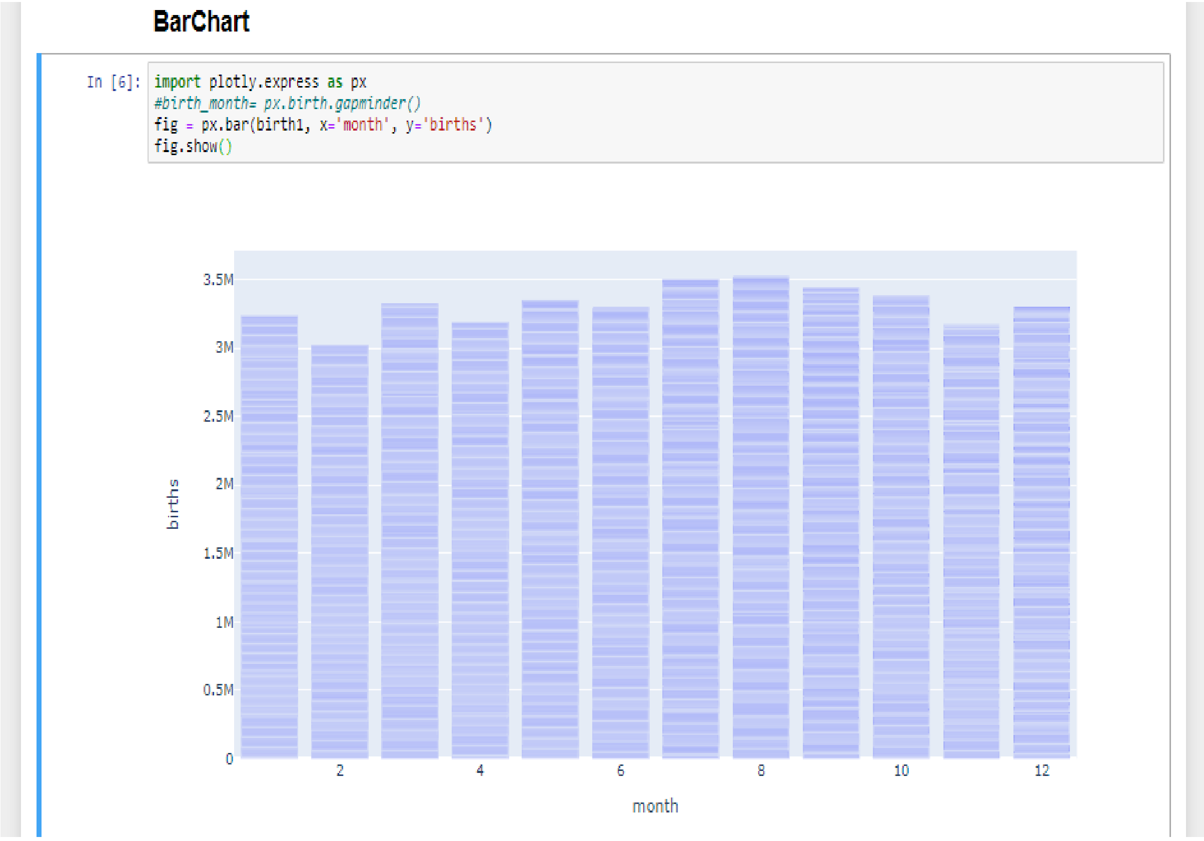
JUPYTER NOTEBOOK:



```
In [2]: import numpy as np
        import pandas as pd
        import seaborn as sns
        import matplotlib as plt
        from sklearn.preprocessing import LabelEncoder
        from sklearn.impute import SimpleImputer
        import mglearn
        import plotly
        # connected=True means it will download the latest version of plotly javascript library.
        plotly.offline.init_notebook_mode(connected=True)
        import plotly.graph_objs as go
        import plotly.figure_factory as ff
        import cufflinks as cf
        import warnings
        warnings.filterwarnings('ignore')

In [4]: birth1 = pd.read_csv(r'C:\Users\tejaswi\Downloads\dv\US_births_1994-2003_CDC_NCHS.csv')
```

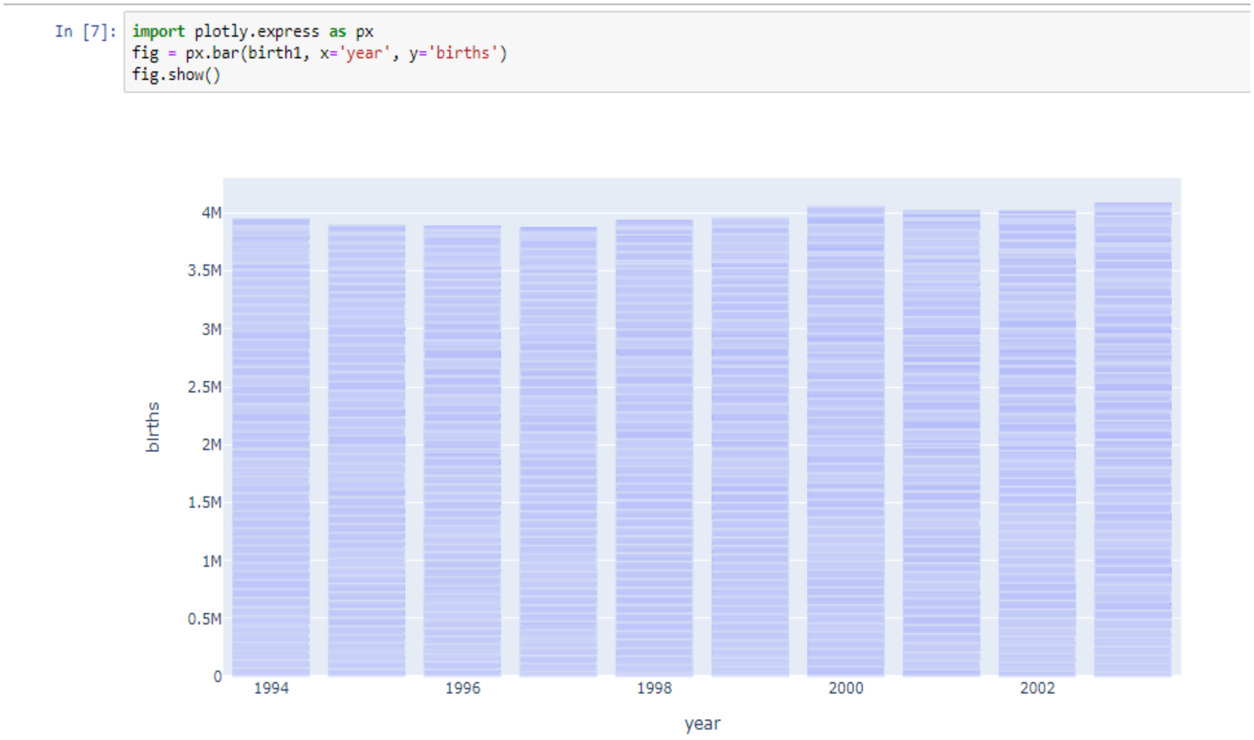


```
In [5]: birth1.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 3652 entries, 0 to 3651
Data columns (total 5 columns):
year             3652 non-null int64
month            3652 non-null int64
date_of_month    3652 non-null int64
day_of_week      3652 non-null int64
births           3652 non-null int64
dtypes: int64(5)
memory usage: 142.8 KB
```

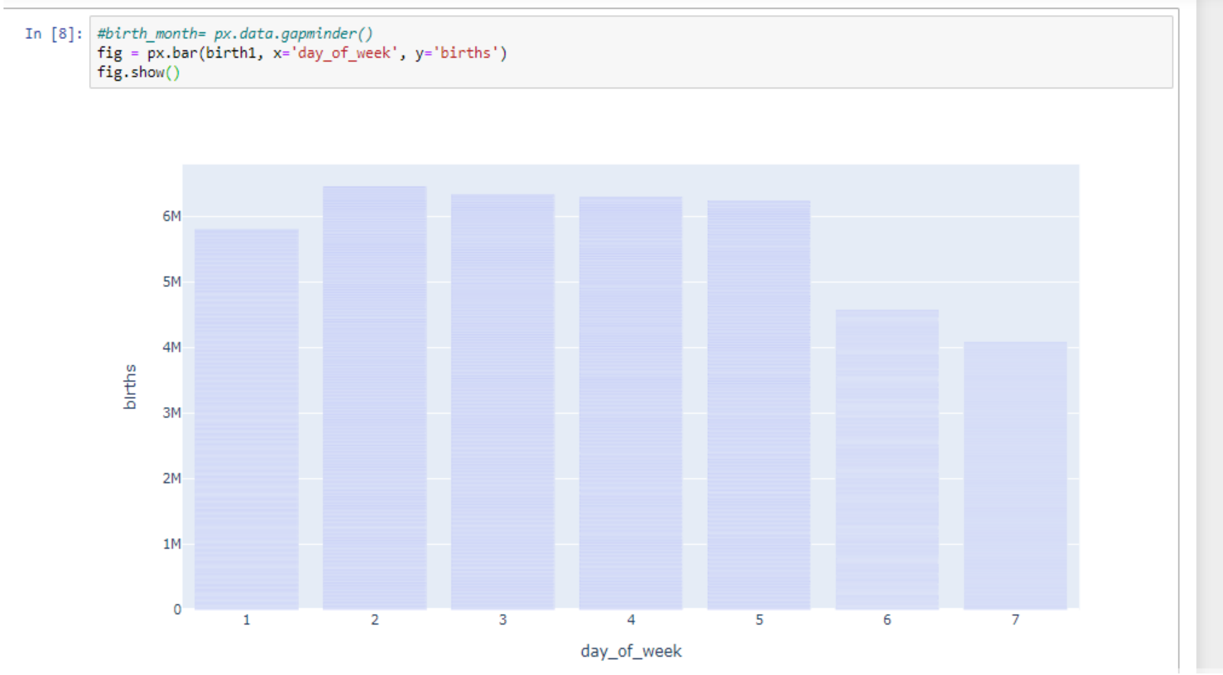


BarChart

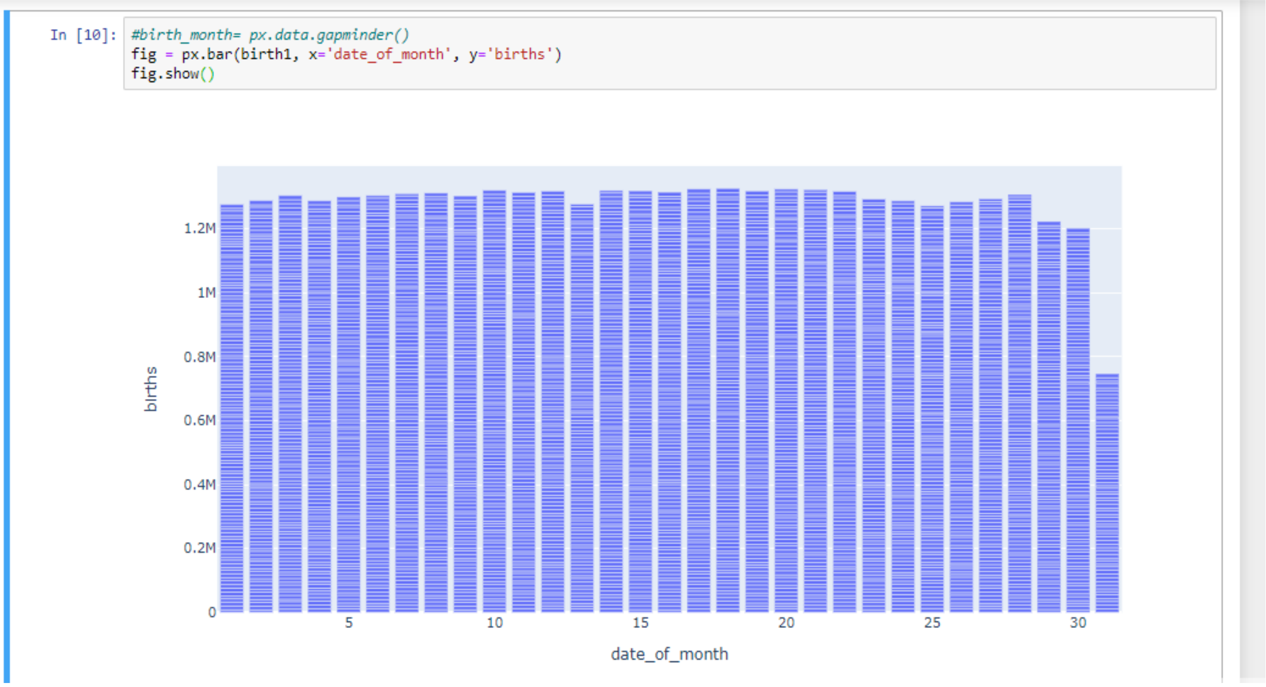
```
In [6]: import plotly.express as px
        #birth_month= px.birth.gapminder()
        fig = px.bar(birth1, x='month', y='births')
        fig.show()
```



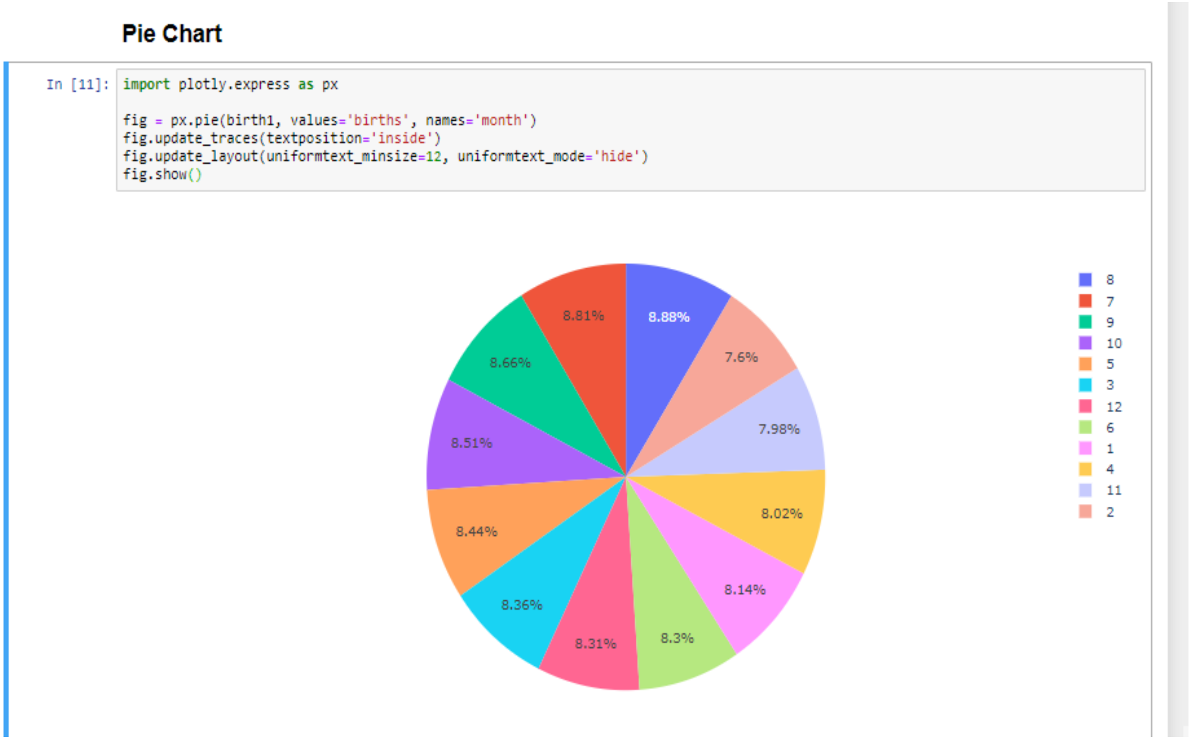
```
In [7]: import plotly.express as px
        fig = px.bar(birth1, x='year', y='births')
        fig.show()
```



```
In [8]: #birth_month= px.data.gapminder()
        fig = px.bar(birth1, x='day_of_week', y='births')
        fig.show()
```

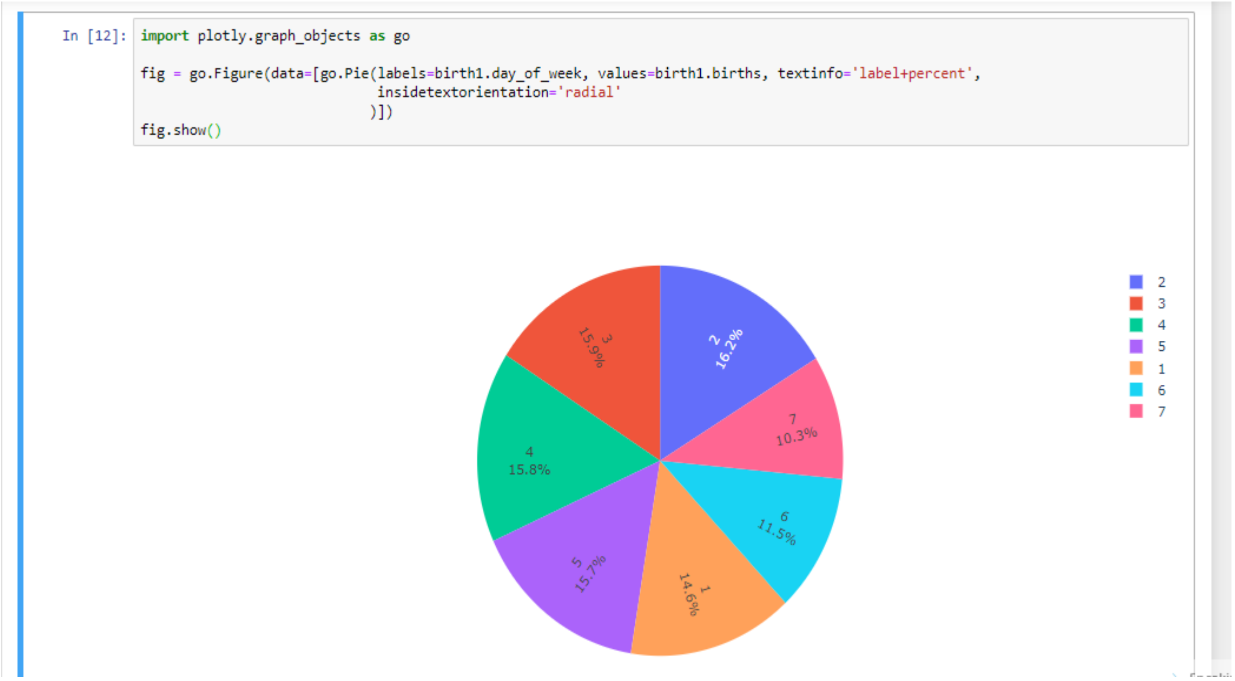


```
In [10]: #birth_month= px.data.gapminder()
         fig = px.bar(birth1, x='date_of_month', y='births')
         fig.show()
```

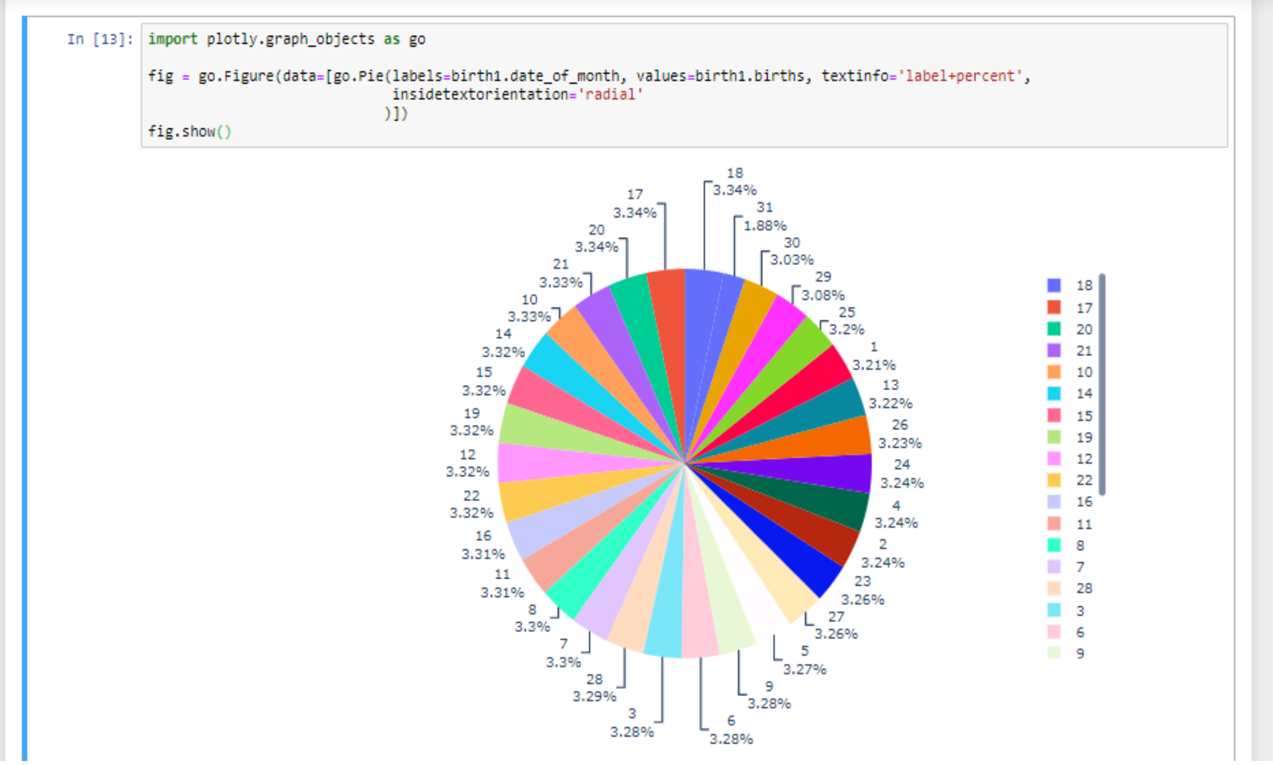


Pie Chart

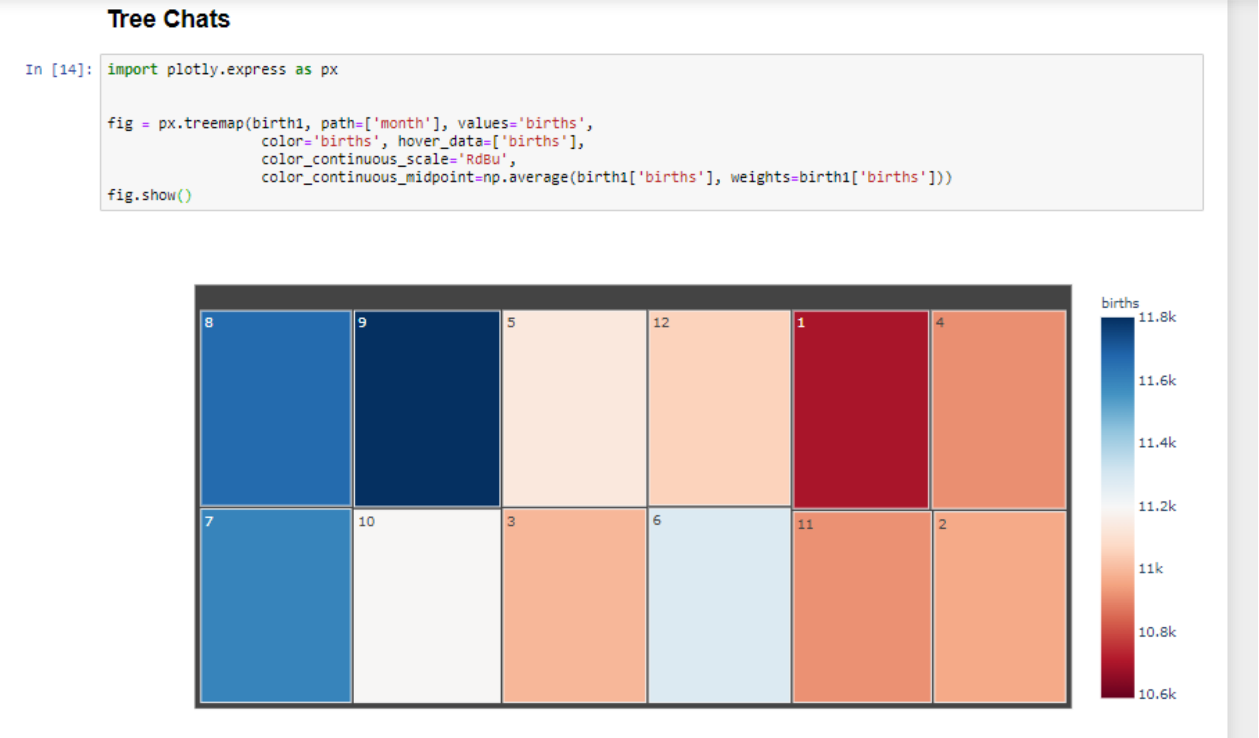
```
In [11]: import plotly.express as px

         fig = px.pie(birth1, values='births', names='month')
         fig.update_traces(textposition='inside')
         fig.update_layout(uniformtext_minsize=12, uniformtext_mode='hide')
         fig.show()
```



```
In [12]: import plotly.graph_objects as go

         fig = go.Figure(data=[go.Pie(labels=birth1.day_of_week, values=birth1.births, textinfo='label+percent',
                                      insidetextorientation='radial'
                                     )])
         fig.show()
```

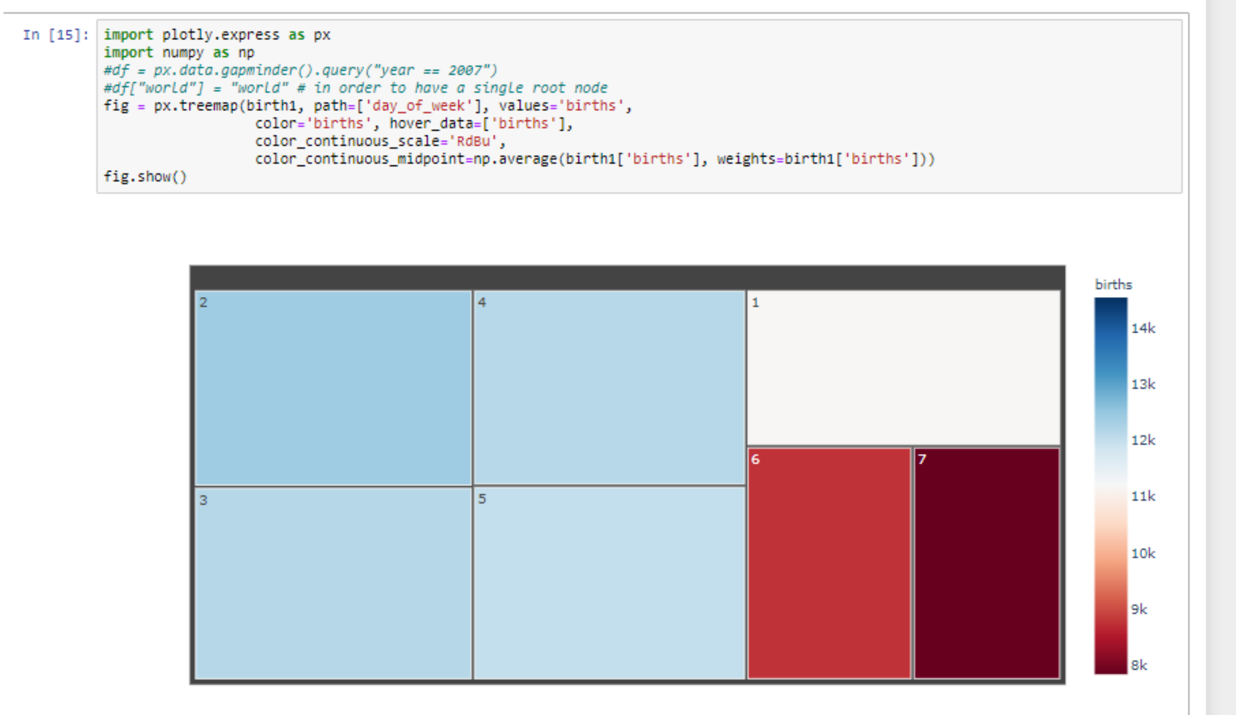


```
In [13]: import plotly.graph_objects as go

         fig = go.Figure(data=[go.Pie(labels=birth1.date_of_month, values=birth1.births, textinfo='label+percent',
                                      insidetextorientation='radial'
                                     )])
         fig.show()
```

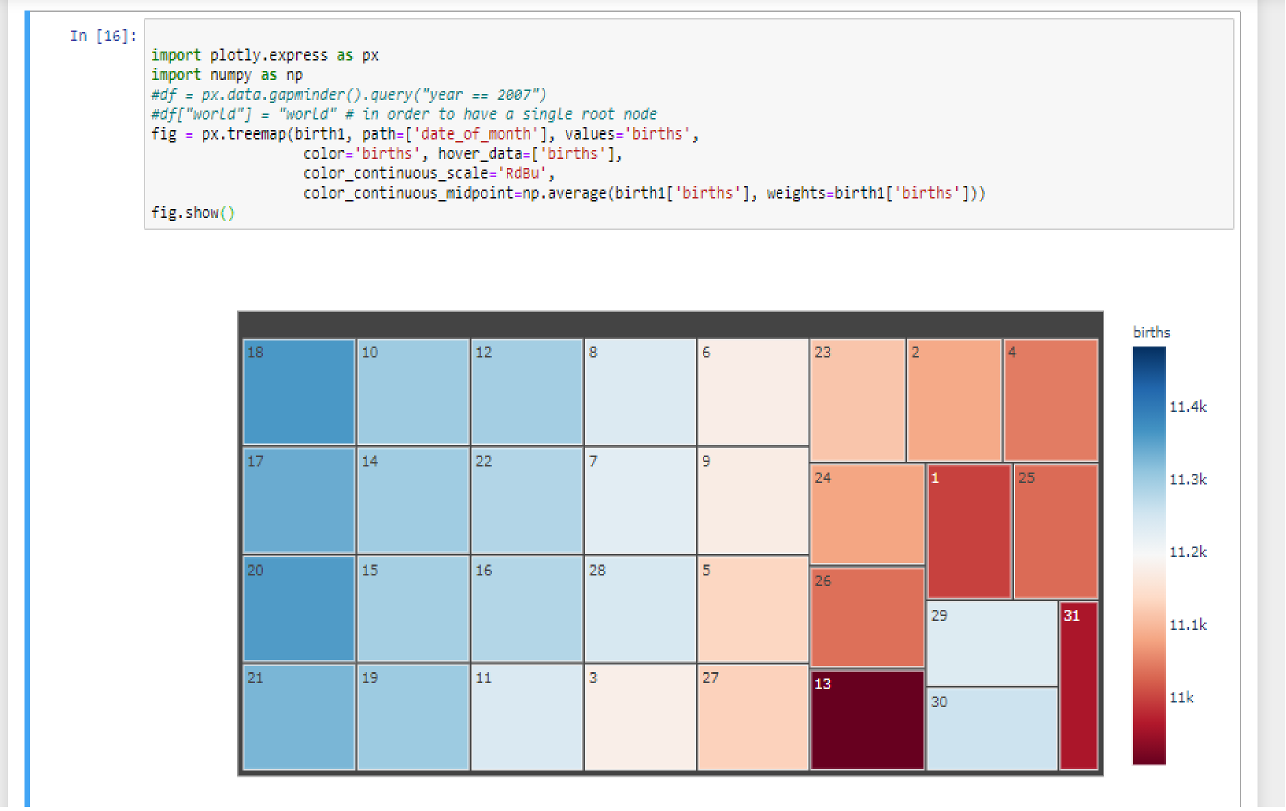


Tree Chats

```
In [14]: import plotly.express as px


         fig = px.treemap(birth1, path=['month'], values='births',
                          color='births', hover_data=['births'],
                          color_continuous_scale='RdBu',
                          color_continuous_midpoint=np.average(birth1['births'], weights=birth1['births']))
         fig.show()
```



In [15]:
```
import plotly.express as px
import numpy as np
#df = px.data.gapminder().query("year == 2007")
#df["world"] = "world" # in order to have a single root node
fig = px.treemap(birth1, path=['day_of_week'], values='births',
                  color='births', hover_data=['births'],
                  color_continuous_scale='RdBu',
                  color_continuous_midpoint=np.average(birth1['births'], weights=birth1['births']))
fig.show()
```



In [16]:
```
import plotly.express as px
import numpy as np
#df = px.data.gapminder().query("year == 2007")
#df["world"] = "world" # in order to have a single root node
fig = px.treemap(birth1, path=['date_of_month'], values='births',
                  color='births', hover_data=['births'],
                  color_continuous_scale='RdBu',
                  color_continuous_midpoint=np.average(birth1['births'], weights=birth1['births']))
fig.show()
```

TABLEAU:

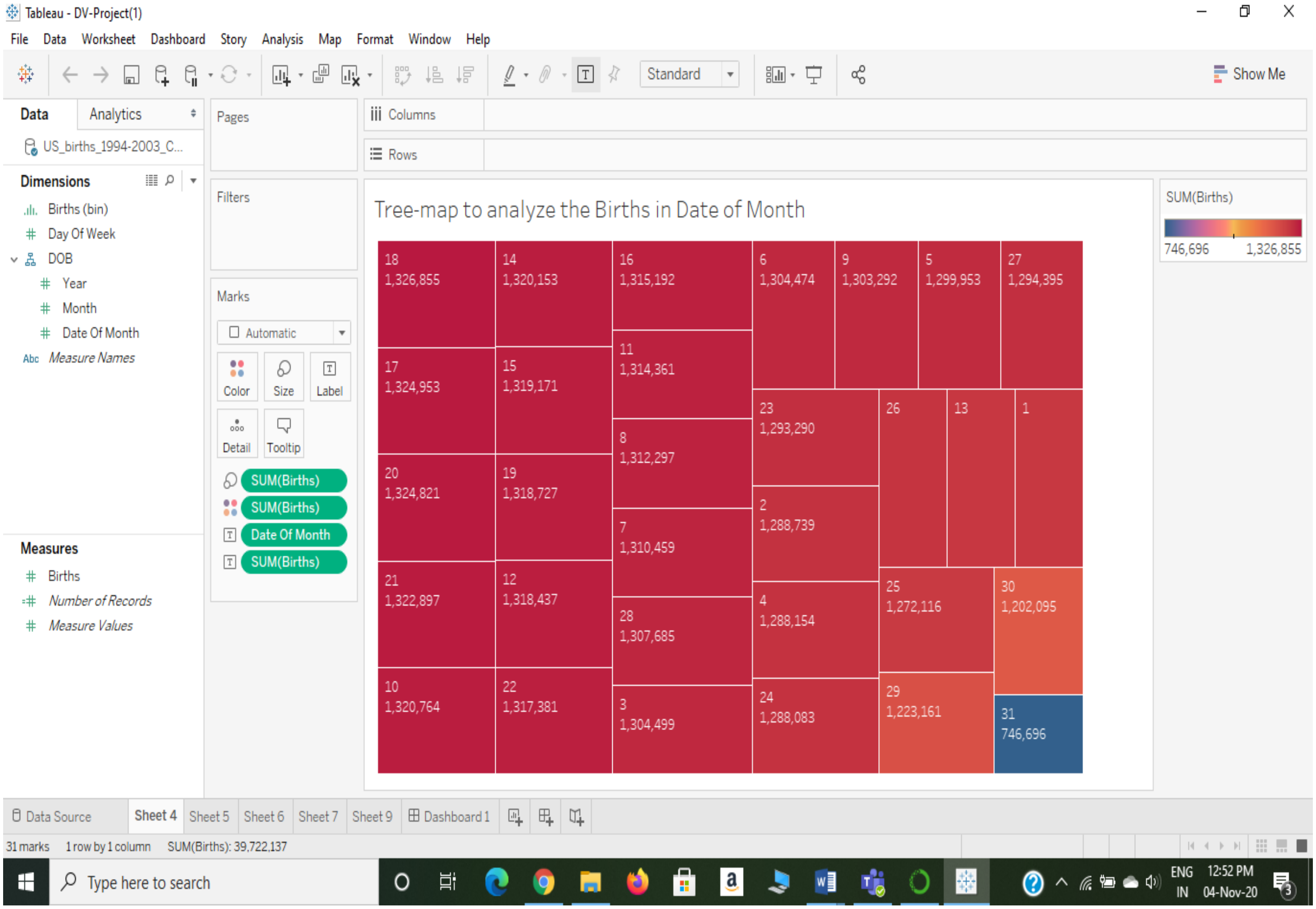


Tableau - DV-Project(1)
File Data Worksheet Dashboard Story Analysis Map Format Window Help
Standard
Show Me
Data
Analytics
US_births_1994-2003_C...
Dimensions
Births (bin)
Day Of Week
DOB
Year
Month
Date Of Month
Measure Names
Measures
Births
Number of Records
Measure Values
Pages
Filters
Marks
Automatic
Color
Size
Label
Detail
Tooltip
SUM(Births)
SUM(Births)
Date Of Month
SUM(Births)
Columns
Rows
Tree-map to analyze the Births in Date of Month
18 1,326,855
14 1,320,153
16 1,315,192
6 1,304,474
9 1,303,292
5 1,299,953
27 1,294,395
17 1,324,953
15 1,319,171
11 1,314,361
23 1,293,290
26
13
1
8 1,312,297
20 1,324,821
19 1,318,727
2 1,288,739
7 1,310,459
21 1,322,897
12 1,318,437
4 1,288,154
25 1,272,116
30 1,202,095
28 1,307,685
10 1,320,764
22 1,317,381
3 1,304,499
24 1,288,083
29 1,223,161
31 746,696
SUM(Births)
746,696
1,326,855
Data Source
Sheet 4
Sheet 5
Sheet 6
Sheet 7
Sheet 9
Dashboard 1
31 marks
1 row by 1 column
SUM(Births): 39,722,137
Type here to search
ENG
IN
12:52 PM
04-Nov-20

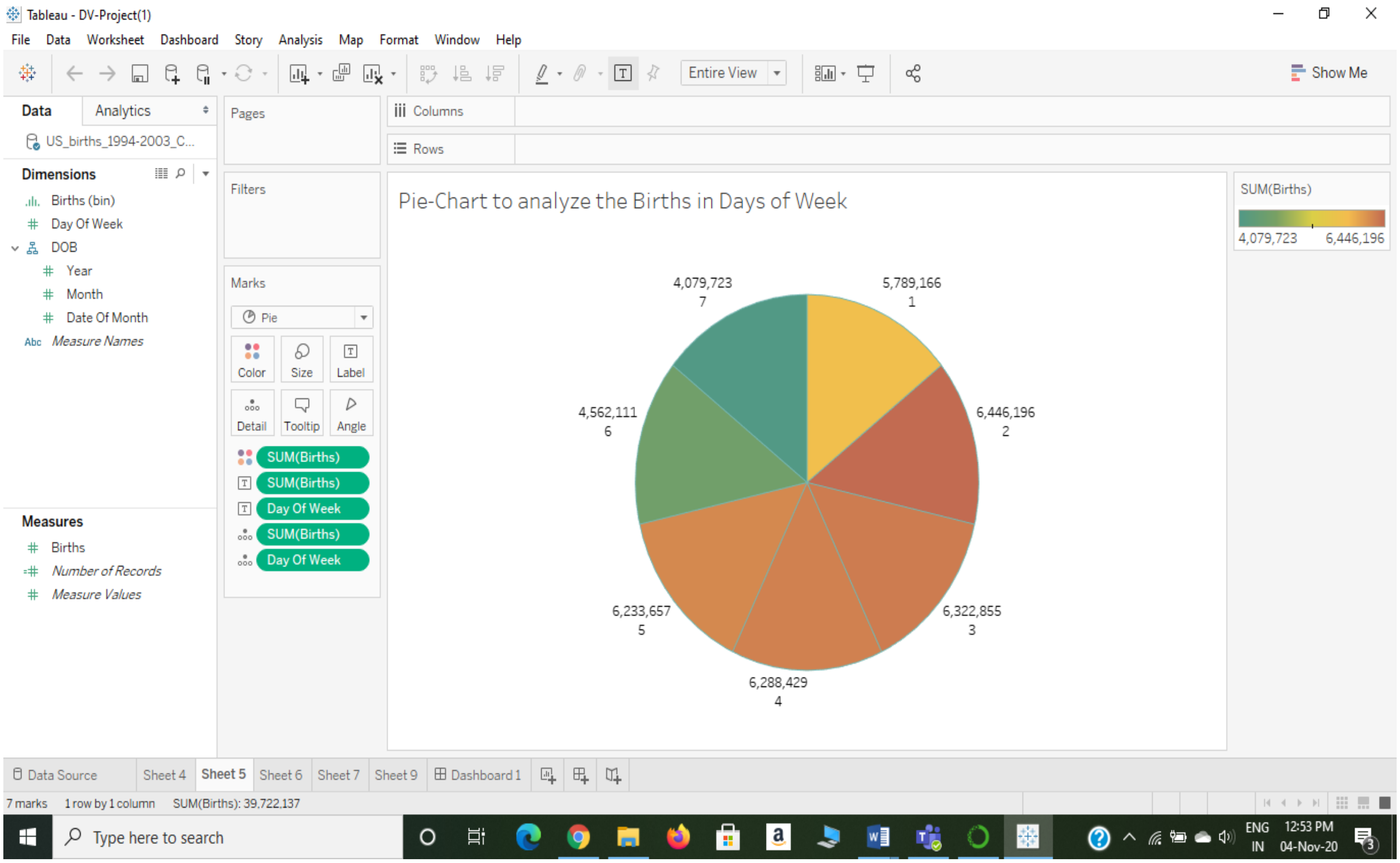


Tableau - DV-Project(1)
File Data Worksheet Dashboard Story Analysis Map Format Window Help
Entire View
Show Me
Data Analytics
US_births_1994-2003_C...
Dimensions
Births (bin)
Day Of Week
DOB
Year
Month
Date Of Month
Measure Names
Measures
Births
Number of Records
Measure Values
Pages
Filters
Marks
Pie
Color Size Label
Detail Tooltip Angle
SUM(Births)
SUM(Births)
Day Of Week
SUM(Births)
Day Of Week
Columns
Rows
Pie-Chart to analyze the Births in Days of Week
4,079,723
7
5,789,166
1
4,562,111
6
6,446,196
2
6,233,657
5
6,322,855
3
6,288,429
4
SUM(Births)
4,079,723
6,446,196
Data Source Sheet 4 Sheet 5 Sheet 6 Sheet 7 Sheet 9 Dashboard 1
7 marks 1 row by 1 column SUM(Births): 39,722,137
Type here to search
ENG IN 12:53 PM 04-Nov-20

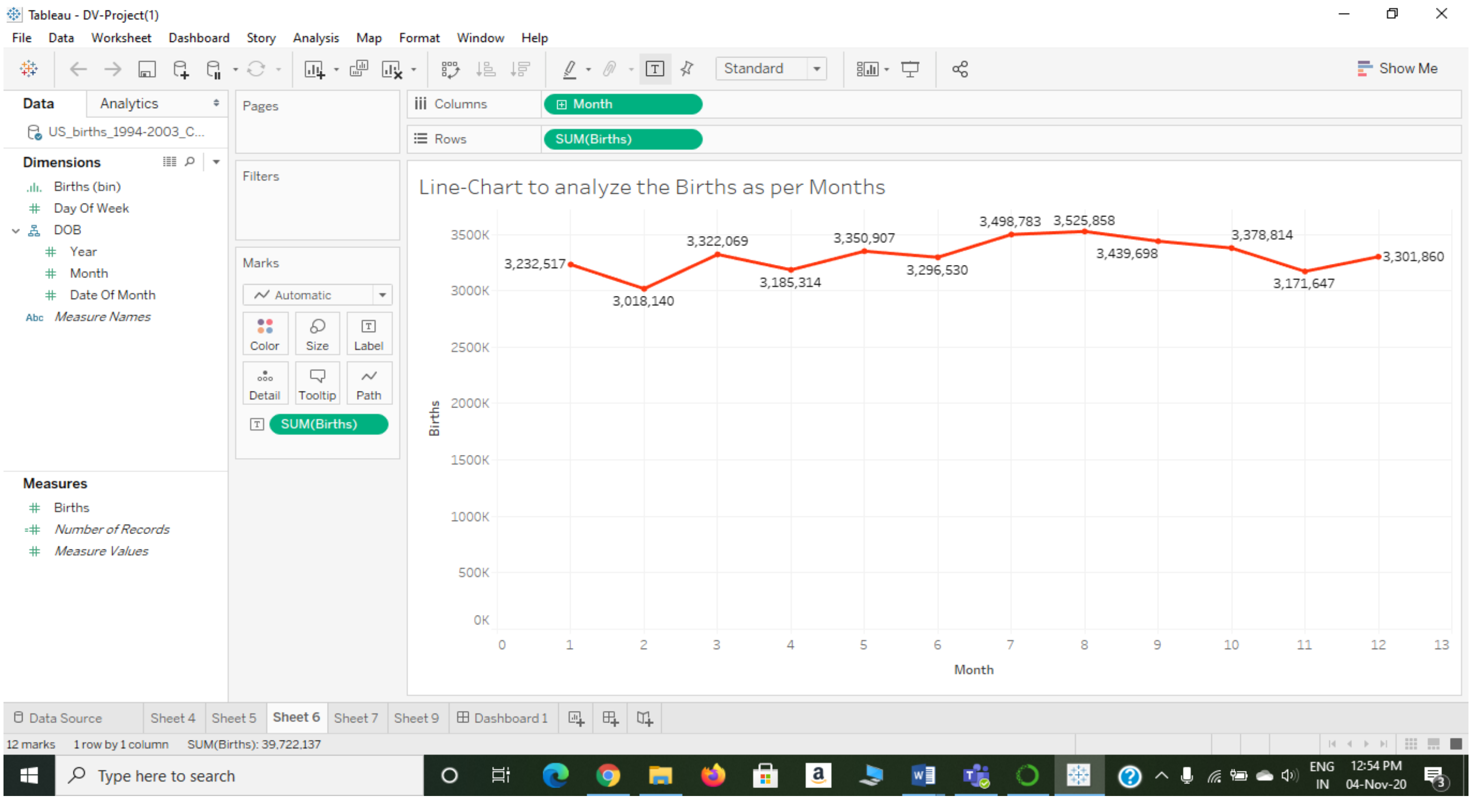


Tableau - DV-Project(1)
File Data Worksheet Dashboard Story Analysis Map Format Window Help
Standard
Show Me
Data Analytics
US_births_1994-2003_C...
Dimensions
Births (bin)
Day Of Week
DOB
Year
Month
Date Of Month
Measure Names
Measures
Births
Number of Records
Measure Values
Pages
Filters
Marks
Automatic
Color Size Label
Detail Tooltip Path
SUM(Births)
Columns Month
Rows SUM(Births)
Line-Chart to analyze the Births as per Months
3,232,517
3,018,140
3,322,069
3,185,314
3,350,907
3,296,530
3,498,783
3,525,858
3,439,698
3,378,814
3,171,647
3,301,860
3500K
3000K
2500K
2000K
1500K
1000K
500K
0K
Births
0 1 2 3 4 5 6 7 8 9 10 11 12 13
Month
Data Source Sheet 4 Sheet 5 Sheet 6 Sheet 7 Sheet 9 Dashboard 1
12 marks 1 row by 1 column SUM(Births): 39,722,137
Type here to search
ENG IN 12:54 PM 04-Nov-20

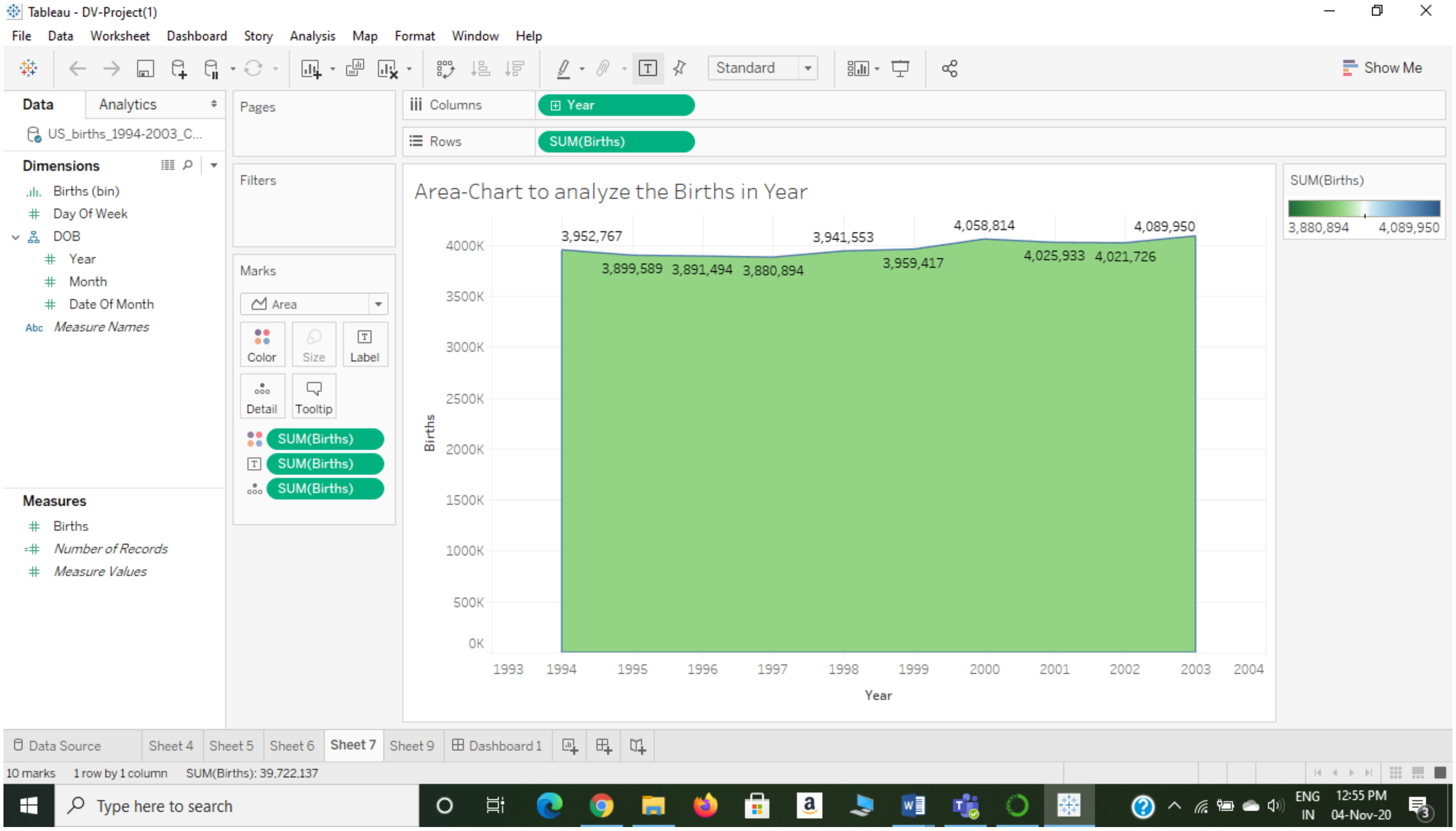


Tableau - DV-Project(1)

File Data Worksheet Dashboard Story Analysis Map Format Window Help

Data: US_births_1994-2003_C...

Columns: Year

Rows: SUM(Births)

Marks: Area — SUM(Births) (Color), SUM(Births) (Label), SUM(Births) (Detail)

Sheet 7

10 marks 1 row by 1 column SUM(Births): 39,722,137

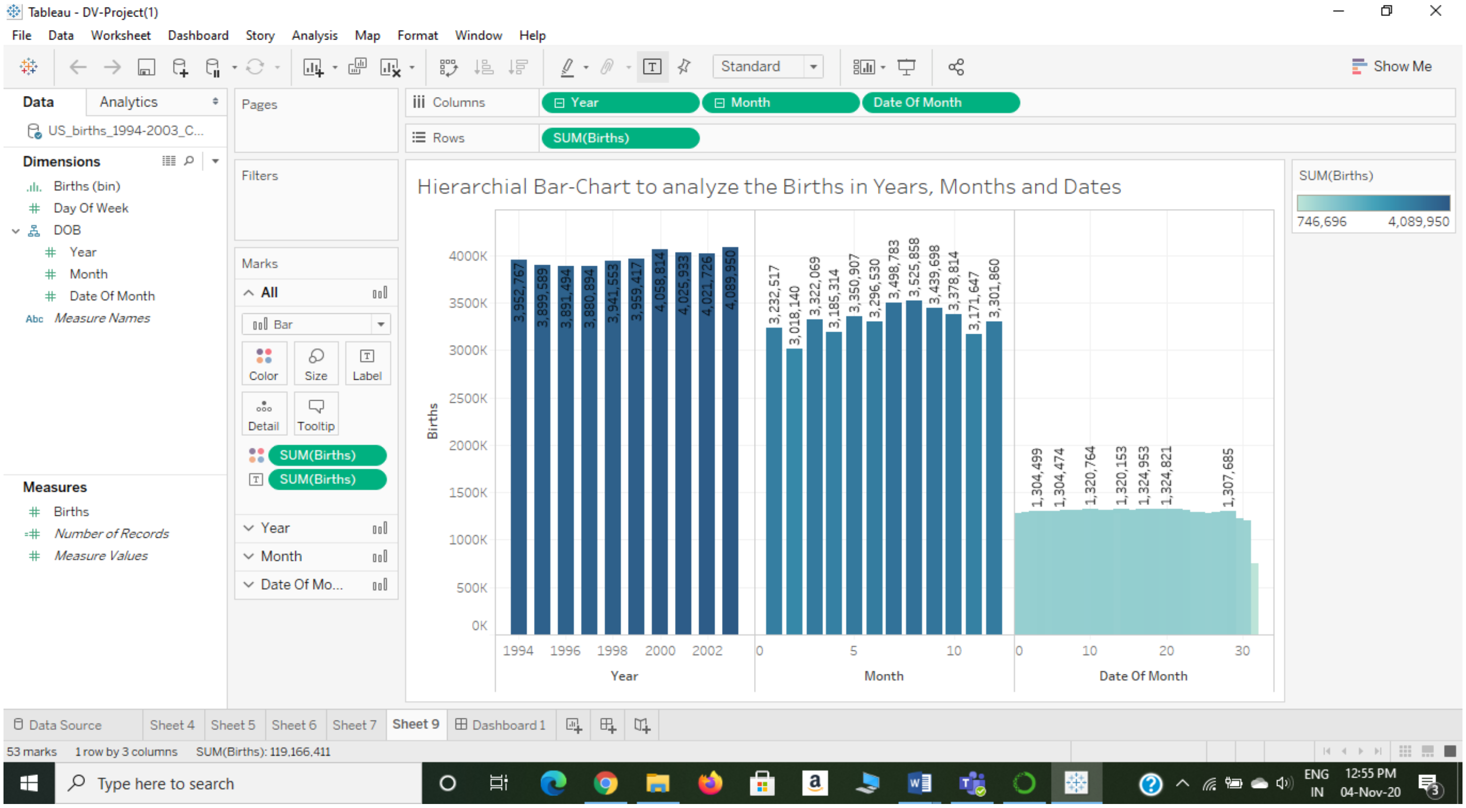


Tableau - DV-Project(1)

File Data Worksheet Dashboard Story Analysis Map Format Window Help

Data: US_births_1994-2003_C...

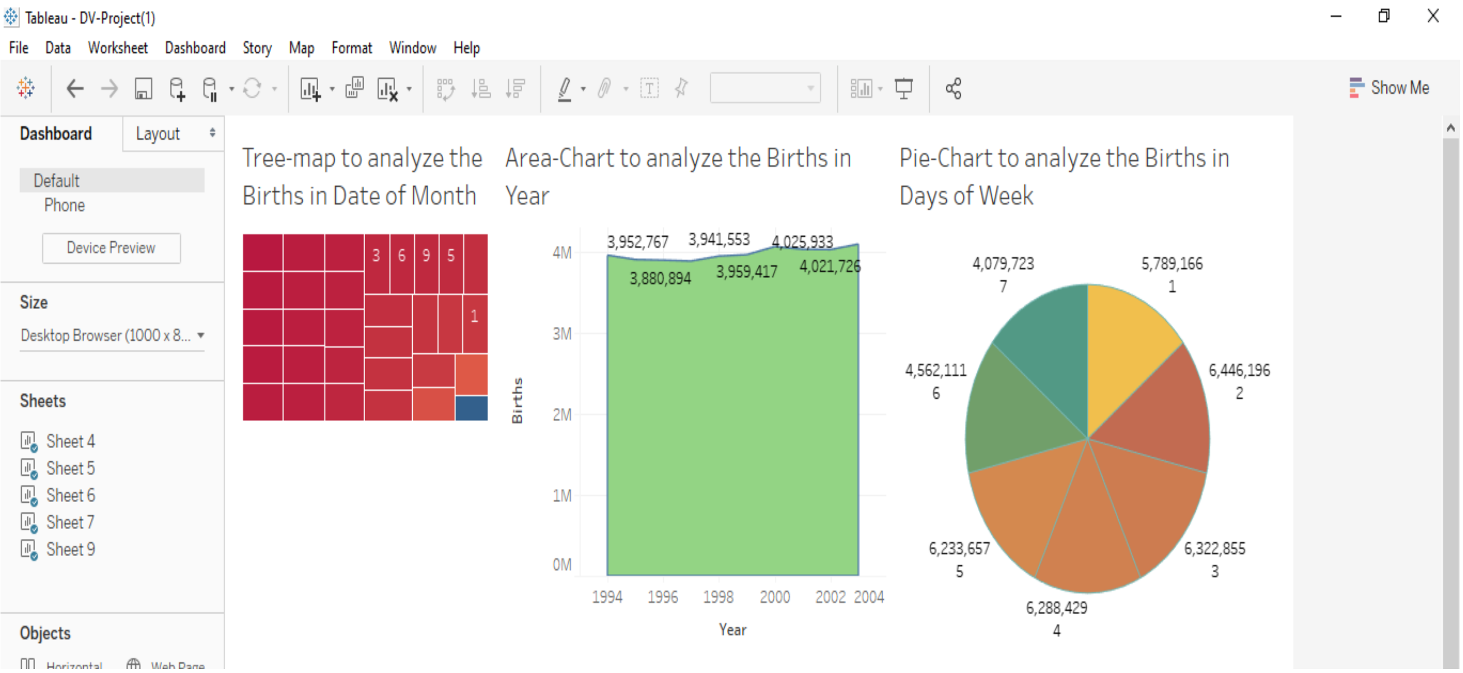
Columns: Year, Month, Date Of Month

Rows: SUM(Births)

Marks: All — Bar — SUM(Births) (Color), SUM(Births) (Label); Year, Month, Date Of Mo...

Sheet 9

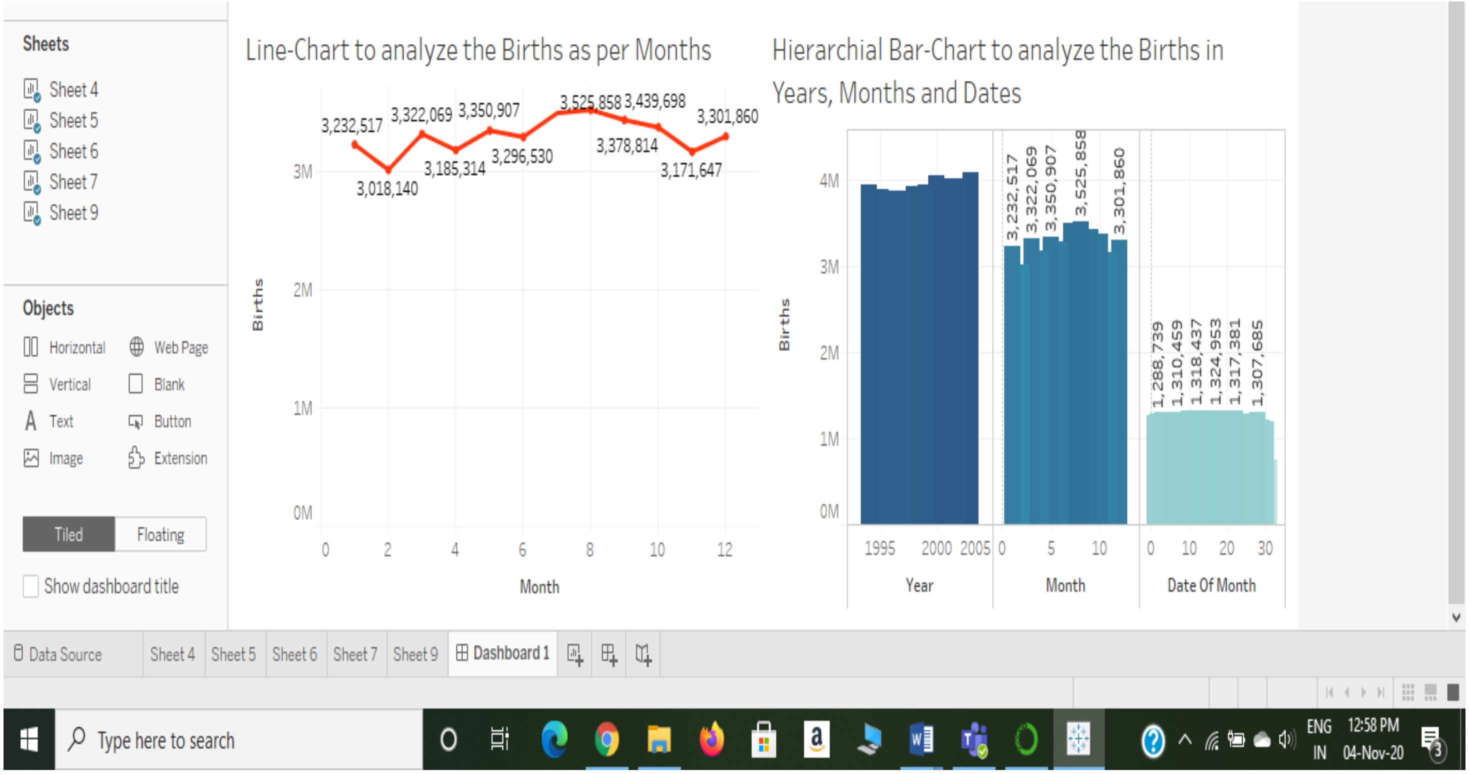
53 marks 1 row by 3 columns SUM(Births): 119,166,411



Tree-map to analyze the Births in Date of Month

Area-Chart to analyze the Births in Year

Pie-Chart to analyze the Births in Days of Week



Line-Chart to analyze the Births as per Months

Hierarchial Bar-Chart to analyze the Births in Years, Months and Dates

CONCLUSION

Through this project, we found the change in birth fertility from the year 1994 to the year 2003. Here we found that there are less births taking place in 31st of the month and maximum births taking place on Tuesdays' of the week and most of the babies are born in the months of July, August and September. In this way we can make many visualizations for a given dataset using Tableau and Python.

REFERENCES

Hedegaard H, Miniño AM, Warner M. Drug Overdose Deaths in the United States, 1999–2018.pdf icon NCHS Data Brief, no 356. Hyattsville, MD: National Center for Health Statistics. 2020.

Wilson N, Kariisa M, Seth P, et al. Drug and Opioid-Involved Overdose Deaths—United States, 2017-2018. MMWR Morb Mortal Wkly Rep 2020;69:290-297.