KARNATAK LAW SOCIETY’S

**GOGTE INSTITUTE OF TECHNOLOGY**

UDYAMBAG, BELAGAVI-590008

**DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING**

A COURSE ACTIVITY REPORT ON

**“DATA VISUALIZATION”**

Submitted for the requirements of 5th semester B.E. in CSE for

**“DATA MINING AND DATA WAREHOUSE”**

Submitted by

|  |  |
| --- | --- |
| **NAME** | **USN** |
| SATWIK JOSHI | 2GI19CS136 |
| ANIKETH MAHADIK | 2GI19CS019 |
| DARSHAN V PATIL | 2GI19CS037 |
| VIJAY HOSURKAR | 2GI20CS420 |

Under the guidance of

Prof.Namitha Bhat

Asst. Prof., Dept. of CSE

Academic Year 2021-2022 (Odd semester)

**Certificate**

Karnatak Law Society’s

GOGTE INSTITUTE OF TECHNOLOGY

UdyambagBelagavi -590008

Karnataka, India.

**Department of Computer Science and Engineering**



This is to certify that the Project work titled “**DATA VISUALIZATION**” carried out by **SATWIK J, ANIKETH M, DARSHAN PATIL, VIJAY HOSURKAR** bearing **USNs:2GI19CS136,2GI19CS019,2GI19CS037,2GI20CS420 submitted** in partial fulfilment of the requirements for 5th semester B.E. in COMPUTER SCIENCE AND ENGINEERING, Visvesvaraya Technological University, Belagavi. It is certified that all corrections/suggestions indicated have been incorporated in the report. The project report has been approved as it satisfies the academic requirements in respect of research work prescribed for the said degree.

**Date: Signature of Guide**

**Place:**Belagavi Prof.Namitha Bhat

**Asst.Prof in Dept. of CSE**

KLS Gogte Institute of Technology,

Belagavi

**Team Members Details:**

|  |  |  |
| --- | --- | --- |
| **S. No.** | **USN** | **Student Name** |
| 1 | **2GI19CS136** | **SATWIK JOSHI** |
| 2 | **2GI19CS019** | **ANIKETH MAHADIK** |
| 3 | **2GI19CS037** | **DARSHAN V PATIL** |
| 4 | **2GI20CS420** | **VIJAY HOSURKAR** |

**Marks Allocation:**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | Batch No. : 03 | | | | | |
| 1. | Project Title:Data Visualization | Marks Range | USN | | | |
| 2GI19CS136 | 2GI19CS019 | 2GI19CS037 | 2GI20CS420 |
| 2. | Problem statement (PO2) | 0-1 |  |  |  |  |
| 3. | Objectives of Defined Problem statement(PO1,PO2) | 0-2 |  |  |  |  |
| 4. | Design / Algorithm/Flowchart/Methodology(PO3) | 0-3 |  |  |  |  |
| 5. | Implementation details/Function/Procedures/Classes and Objects (Language/Tools)(PO1,PO3,PO4,PO5) | 0-4 |  |  |  |  |
| 6. | Working model of the final solution (PO3,PO12) | 0-5 |  |  |  |  |
| 7. | Report and Oral presentation skill (PO9,PO10) | 0-5 |  |  |  |  |
|  | Total | 20 |  |  |  |  |

**TABLE OF CONTENTS**

* **Introduction**
* **Abstract**
* **Problem Statement**
* **Proposed objectives & methodology**
* **Experimental results**
* **Data sets**
* **Conclusion**

**INTRODUCTION**

Data visualization is the graphical representation of information and data.By using visual elements like charts, graphs, and maps, data visualization tools provide an accessible way to see and understand trends, outliers, and patterns in data.Data visualization is one of the steps of the data sceince process, which states that after data has been collected, processed and modeled, it must be visualized for conclusions to be made.

Data visualization provides a quick and effective way to communicate information in a universal manner using visual information. There are many types of data visualization. The most common are scatterplots, line graphs, pie charts, bar charts, heat maps, area charts, choropleth maps and histograms. Visualization is done by using several libraries in python like – matplotlib, seaborn etc.

Good data visualizations allow us to**reason and think effectively**about our data. By presenting information visually, it allows us offload internal cognition to the perceptual system. If we see numerical data in a table, we may be able to find a trend, but it will take a significant amount of work on our part to recognize and conceptualize that trend. By plotting that data visually, that trend becomes immediately clear to our mind through our perceptual system.

It’s hard to think of a professional industry that doesn’t benefit from [making data more understandable](https://www.forbes.com/sites/jeffkauflin/2017/07/20/the-five-most-in-demand-skills-for-data-analysis-jobs/#3e300312c7ce). Every STEM field benefits from understanding data—and so do fields in government, finance, marketing, history, consumer goods, service industries, education, sports, and so on. While we’ll always wax poetically about data visualization (you’re on the Tableau website, after all) there are practical, real-life applications that are undeniable. And, since visualization is so prolific, it’s also one of the most useful professional skills to develop. The better you can convey your points visually, whether in a dashboard or a slide deck, the better you can leverage that information. The concept of [the citizen data scientist is on the rise](https://www.gartner.com/newsroom/id/3570917). Skill sets are changing to accommodate a data-driven world. It is increasingly valuable for professionals to be able to use data to make decisions and use visuals to tell stories of when data informs the who, what, when, where, and how. While traditional education typically draws a distinct line between creative storytelling and technical analysis, the modern professional world also values those who can cross between the two: data visualization sits right in the middle of analysis and visual storytelling.

**ABSTRACT**

Datasets are the foundation and starting point for visualizing your data. They are defined on the connections to your data and provide access to the specific tables in the data store.

A dataset is the logical representation of the data you want to use to build visuals. It is a logical pointer to a physical table or a defined structure in your data source. Datasets may represent the contents of a single data table or a data matrix from several tables that may be in different data stores on the same connection.

Other than providing access to data, datasets enhance data access and use in many ways, including (but not limited to):

* Table joins allow you to supplement the primary data with information from various other data sources. For more information, see Data modeling.
* Derived fields/attributes support flexible expressions, both for dimensions and for aggregates. For more information, see Creating calculated fields.
* Hiding fields enables you to eliminate the fields that are unnecessary to the business use case or to obscure sensitive data without affecting the base tables. For more information, see Hiding dataset fields from applications.
* Changing data types of the field attributes often helps you to deal with data types, or to ensure that numeric codes (like event ids) are processed correctly. For more information, see Changing data type.
* Changing the default aggregation of fields at the dataset level prevents common mistakes when building visuals. For more information, see Changing field aggregation.
* Providing user-friendly names for native columns or derived attributes often makes the visuals more accessible and saves some of the efforts of applying aliases to each field of the visual. For more information, see Automatically renaming dataset fields and Custom renaming dataset fields.
* Different datasets are created in different ways. In this post, you’ll find links to sources with all kinds of datasets. Some of them will be machine-generated data. Some will be data that’s been collected via surveys. Some may be data that’s recorded from human observations. Some may be data that’s been scraped from websites or pulled via APIs.

**PROBLEM STATEMENT**

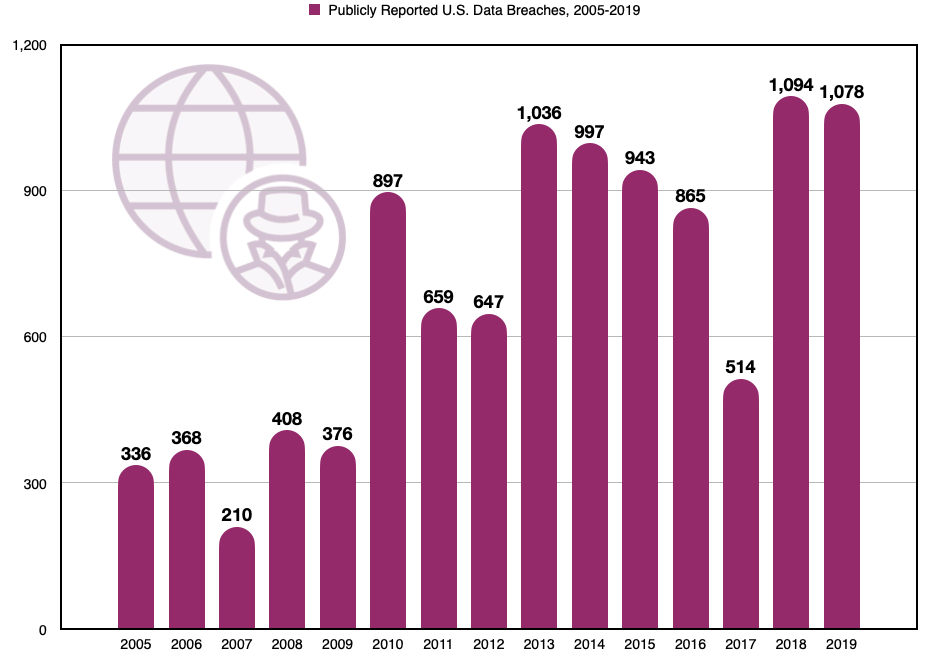
The Data set given contains the data for flights and its detailed services . Analyze the data set and its details , And represent the data set as organized form like bar charts , histogram , pie chart etc. Using python in arranged format with relevant attributes and names.

Thanks to visualization, most complex problems can be broken down into simpler elements for data scientists to figure out the optimal model architectures and solutions to complicated tasks. Hence, visualizations play a vital role in the successful completion of every major Data Science project. Without the use of visualization, it is nearly impossible to gauge the data patterns of a difficult task.

Here we will understand some of the basic features of data visualizations and try to interpret the benefits of exploratory data analysis while solving any kind of task. We will discuss a few essential libraries in python for visualization purposes . Then, we will have a detailed solution on the problem . Finally, we will conclude with a real-time example of interpreting these visualizations.

**Proposed Objectives & Methodology**:

For example, a bar chart communicates data breach trends more effectively than the agent-provided chart. While the data story may not be as dramatic as the non-objective data breach chart, you'll build better, longer-lasting relationships with stakeholders by stating the facts and letting the target audience decide the issue. Spread your research across multiple groups representative of your target audience. This includes using internal tools and working with outside vendors to gain objective viewpoints about the data.



The artistic representation of data changes as it passes through a number of phases, and over the time, its objective also changes with the up gradation of analysis and presentation. Historical examples will show that, many data visualization was incomplete which failed to serve the purpose. **But as time passed, the perception on data visualization changed while making a**[multipurpose decision](https://www.ernestoolivares.com/we-are-90-visuals-beings/)**. It’s changeable nature frequently satisfied various optimizations with proper reasoning.** For comprehension, it became a necessity to look deeply at the base of the work that highlights the creator’s main goal and ideas. Our ability is limited without a goal, and we will not be able to satisfy our customer related to different fields. So as a designer it is our challenge to be able to combine data, extract and transfer information in a single unit to the decision-makers in a sensible and useful form.

One of the main goals of data visualization is to give support in making decision through appropriately designed graphically represented information. Sophistically-designed data visualization systems can greatly assist users with proper reasoning and decision making. Usually, visually presented data is easier to understand.

A visually formatted information allows the user to swiftly perceive patterns or features that may not have been discovered previously.

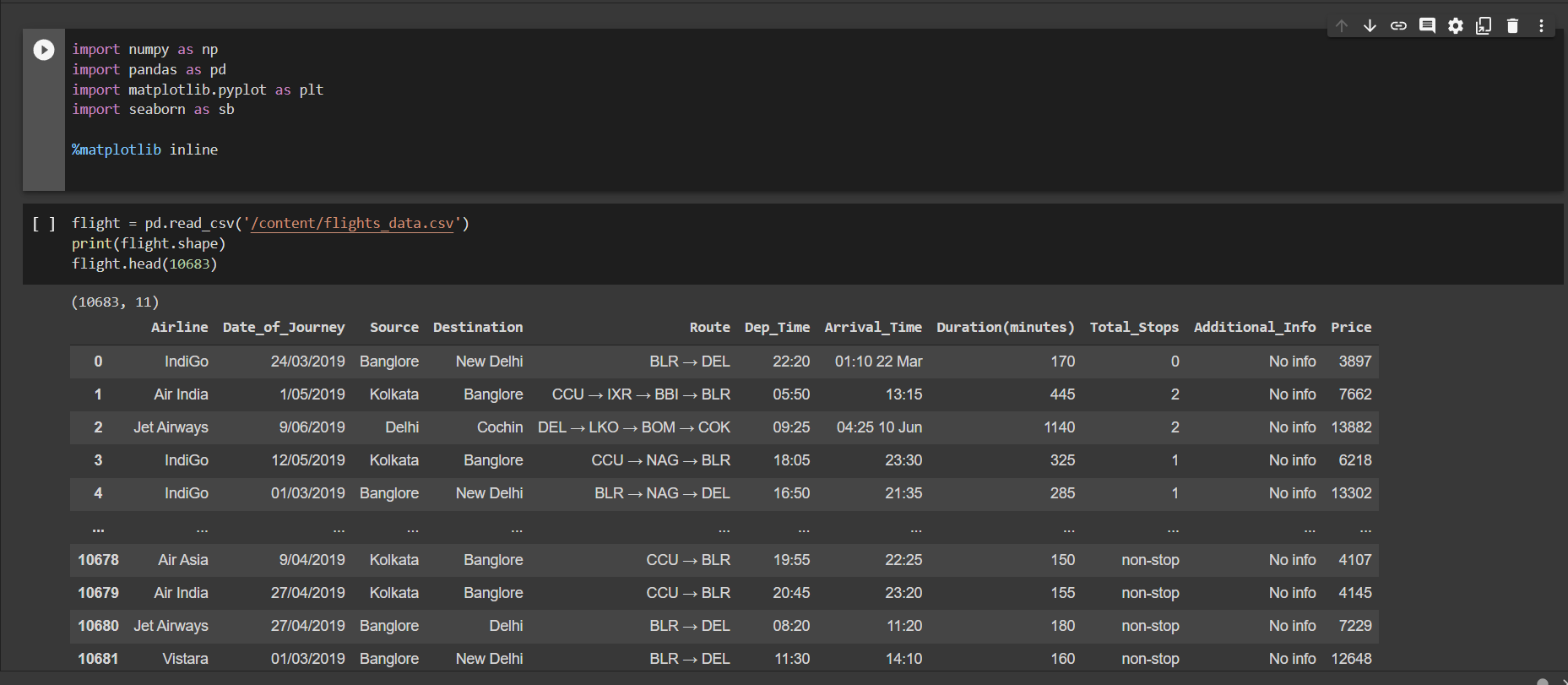
With rapid evolution, the goal of data visualization has moved from simple charts and graphs to strong visually appealing dashboards that everybody wants. However, most of the time it fails to reach its ultimate goal due to its lack of communication which gets lost in unnecessary detail.

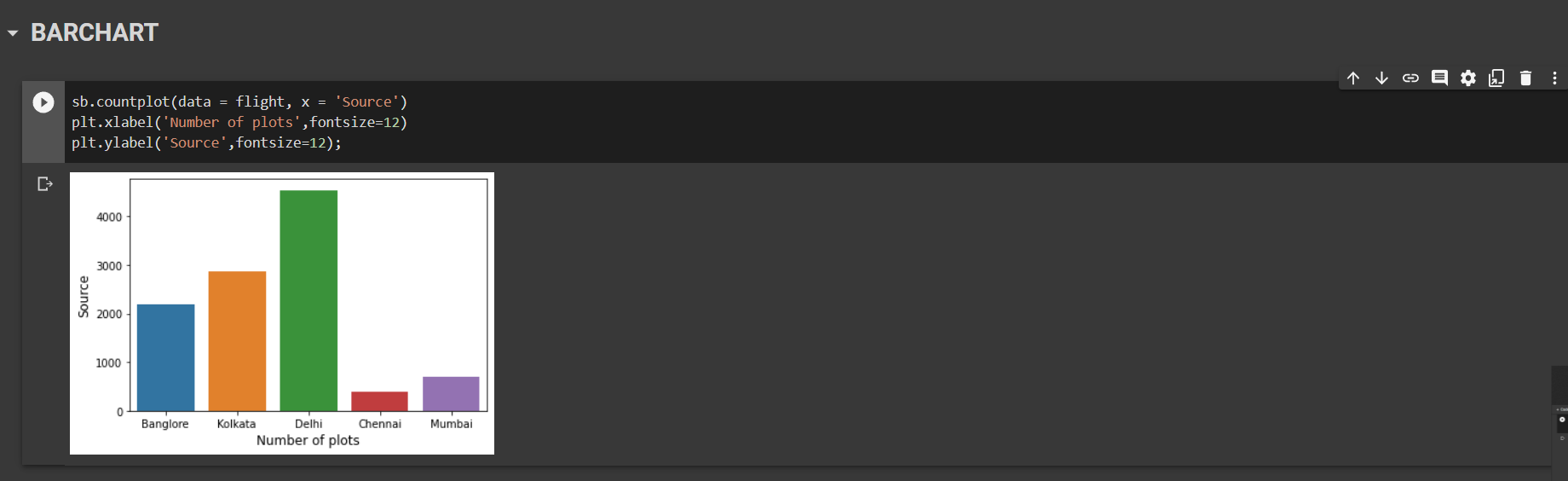
To achieve the goal of data visualization, an organization should properly and efficiently implement a dashboard so as to facilitate instant understanding. This would enhance the capability of working faster and smarter.

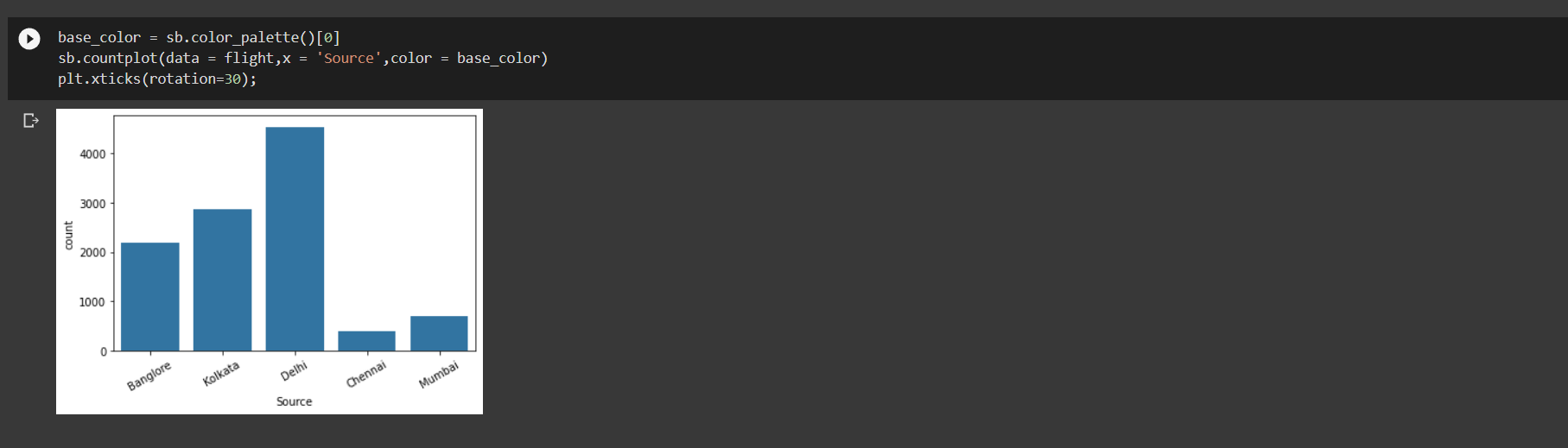
The given data set, using python and some important libraries like matplotlib , numpy etc. we analyse the dataset and represent it in many forms like bar graphs,pie chart , histogram etc.

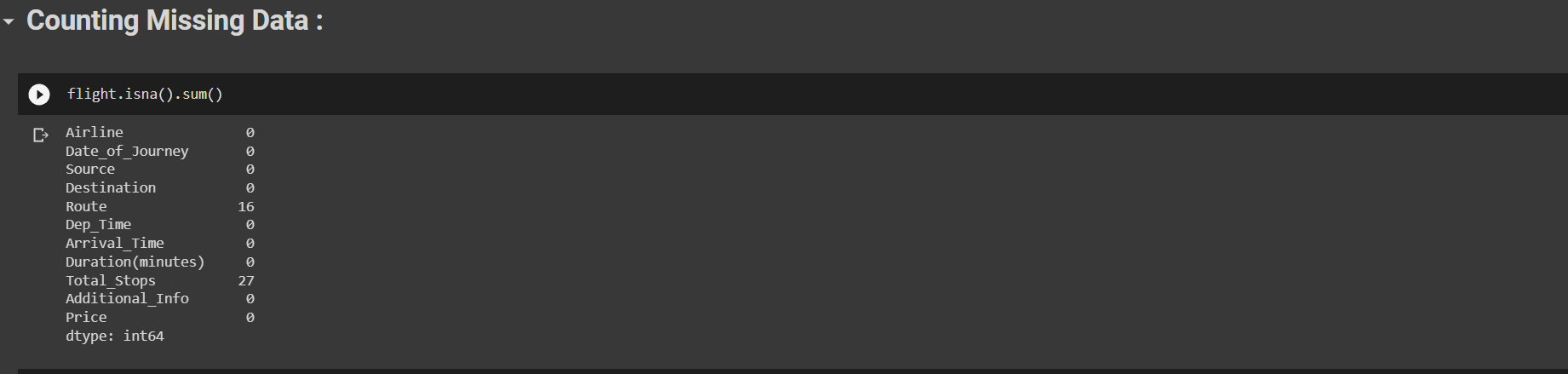
**Experimental Results** :

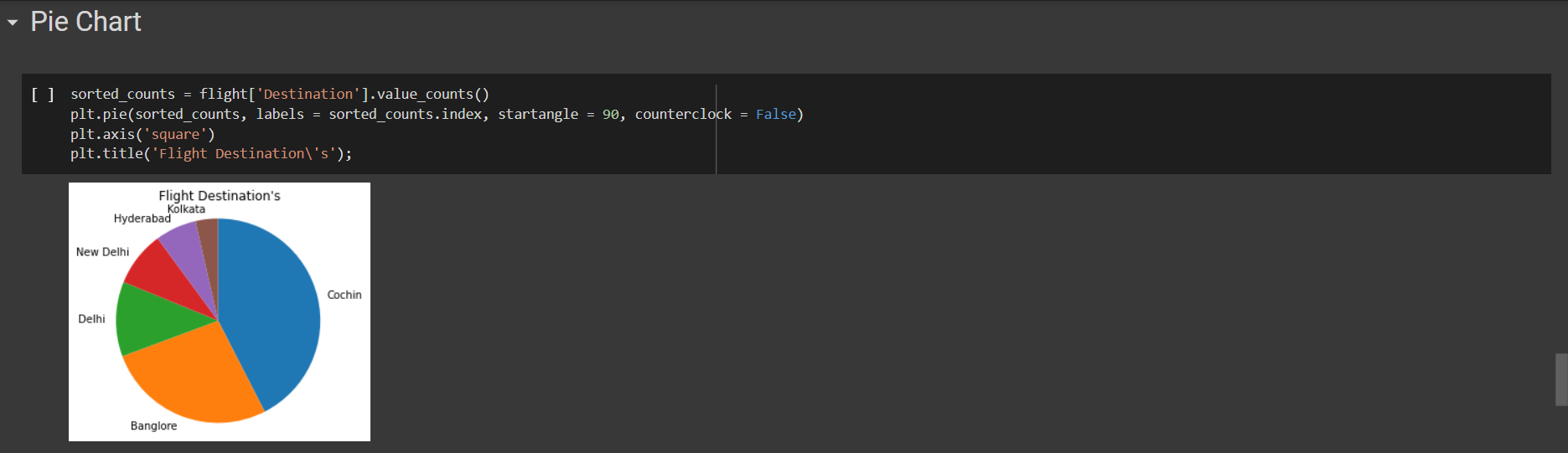
Implemention:



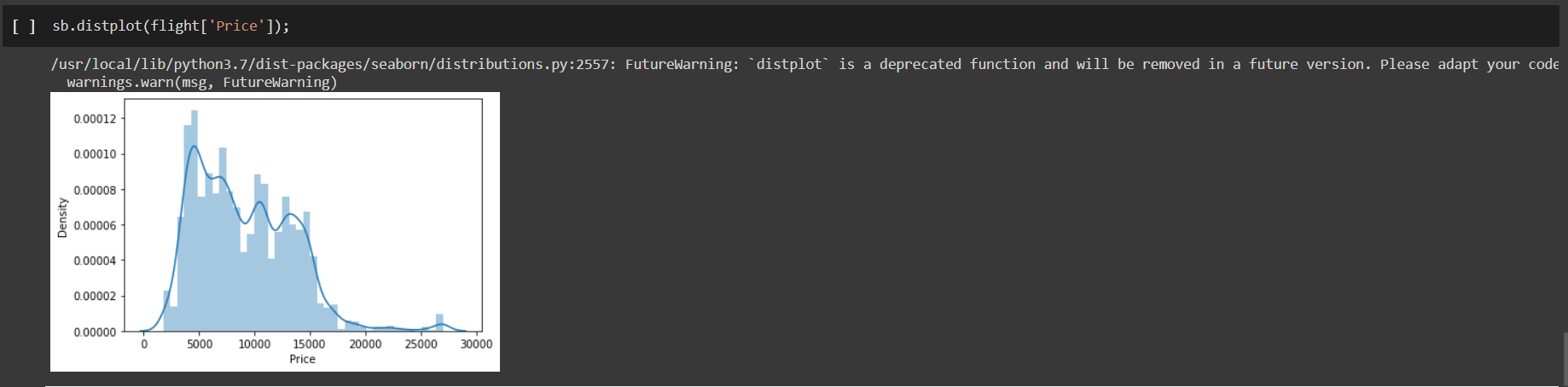


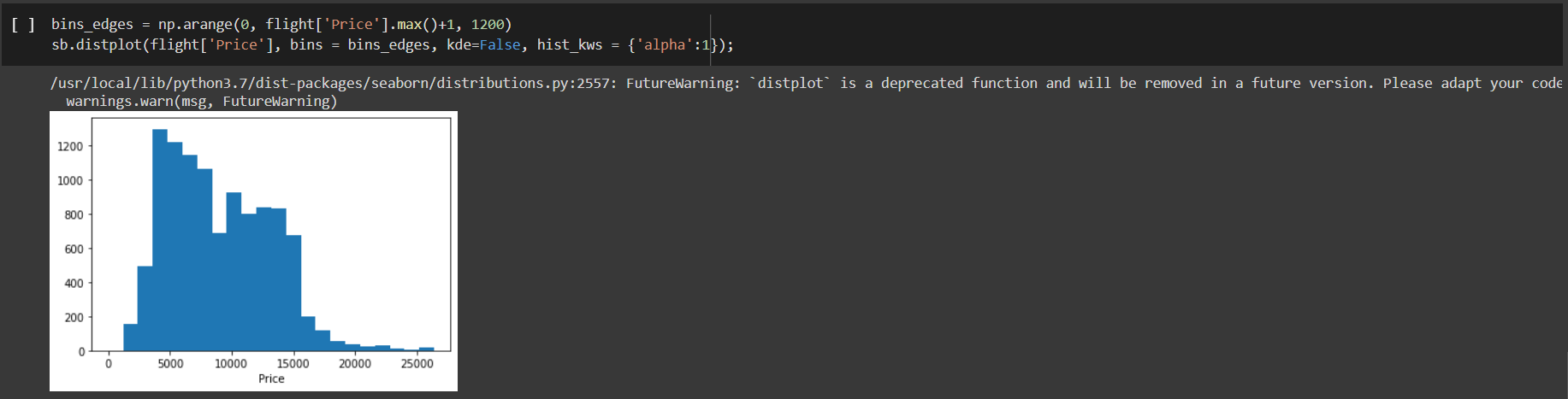




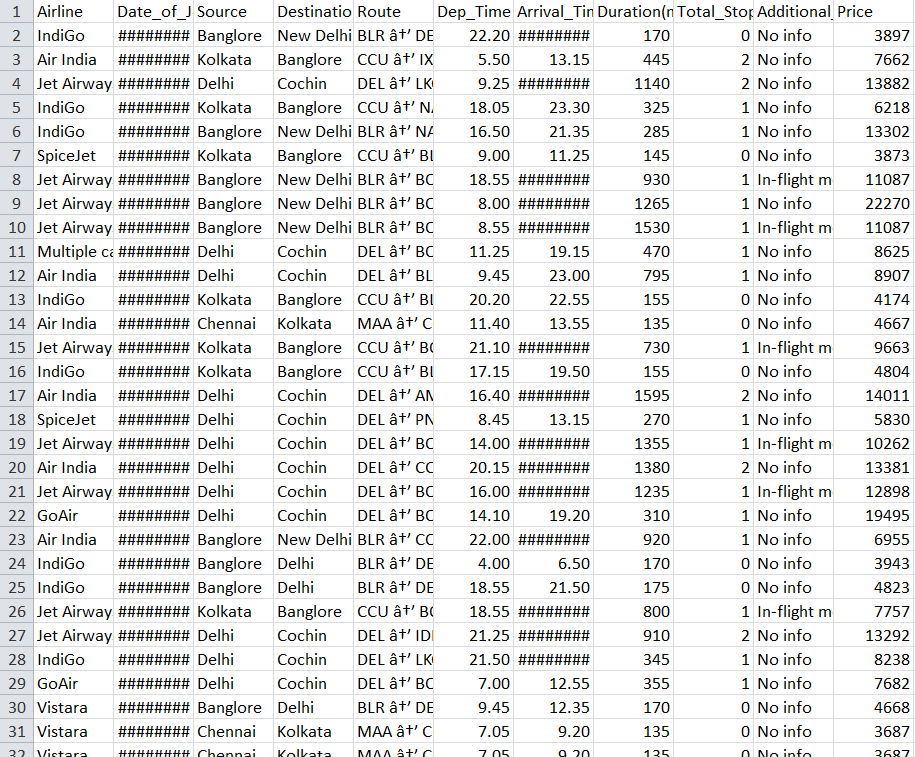


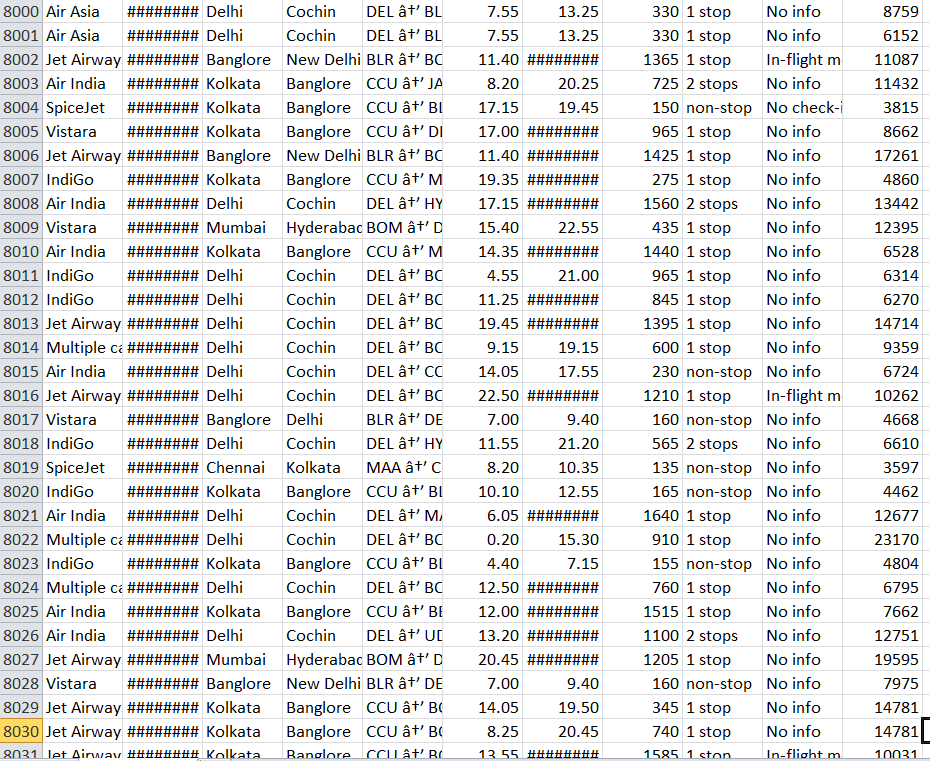






Data Set (The details of flights):





The dataset used for experimentation was retrieved from “kaggle.com”. The chosen dataset has 10683 rows and 11 key features. The measures used for evaluating the performance are precision, recall, and accuracy. The results are narrowed down to the analysis eliminating the noises and the errors in the data set.

The link of full implementation of the project :

<https://colab.research.google.com/drive/1mHoNkx-EJZhH-tjkRIJNIZQAy16TFbBg?usp=sharing>

**Conclusion**

We need data visualization because a visual summary of information makes it easier to identify patterns and trends than looking through thousands of rows on a spreadsheet. It's the way the human brain works. Charts and graphs make communicating data findings easier even if you can identify the patterns without them.

As we conclude our brief study on data visualization, it is clear that the field is rich in potential applications in diverse disciplines, at the same time we need to be aware of its practical and ethical complexities.previously we implemented the data set in many forms which are easy to comprehend and interpret,using python and important libraries.

**References:**

1.<https://www.tableau.com/learn/articles/data-visualization>.

## 2. [Information Graphics” by Sandra Rendgen, Julius Wiedemann](https://www.amazon.com/Information-Graphics-Sandra-Rendgen/dp/3836528797).

3. <https://www.adamenfroy.com/data-visualization-tools>.

4.https://colab.research.google.com/drive/1mHoNkx-EJZhHtjkRIJNIZQAy16TFbBg?usp=sharing