

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | | | | | | |
|  | | Dataset source Analysis | | |  | |
|  |  | | | | |  |
|  | | |  |  | | |
|  | | | Group C  1. Sayan Mukherjee 2. Ritaban Guchait 3. Anwesha Patra 4. Swagata Nanda 5. Soumik Mukherjee 6. Arindam Pakhira 7. Aritra Mukherjee 8. Ayan Kundu 9. Biswanath Mandal 10. Tamal Mondal |  | | |

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | |  |  | | |  |  | |
|  | | THE PROCESS | | | | |  | |
|  | |  |  | | |  |  | |
|  | Knowing about the data sources We can primarily classify the data sources in two parts, **open source** and **paid.** Open source is data with no copy right and has free redistribution. We have used “**Kaggle**” to find such datasets and their licenses.  Paid data sources are like online stores where you can buy data. These sources offer different types of data, like market research, financial data, or maps, and you need to pay a fee to get access to that data. It's like buying a product, but instead, you're buying valuable information to help you make decisions or learn more about a particular topic. Paid data assures high data integrity, data relevance and low data bias. | | | | | | |  |
|  | Data Licensing | | |  |  | | |  |

Data licenses provide important details about the terms and conditions for

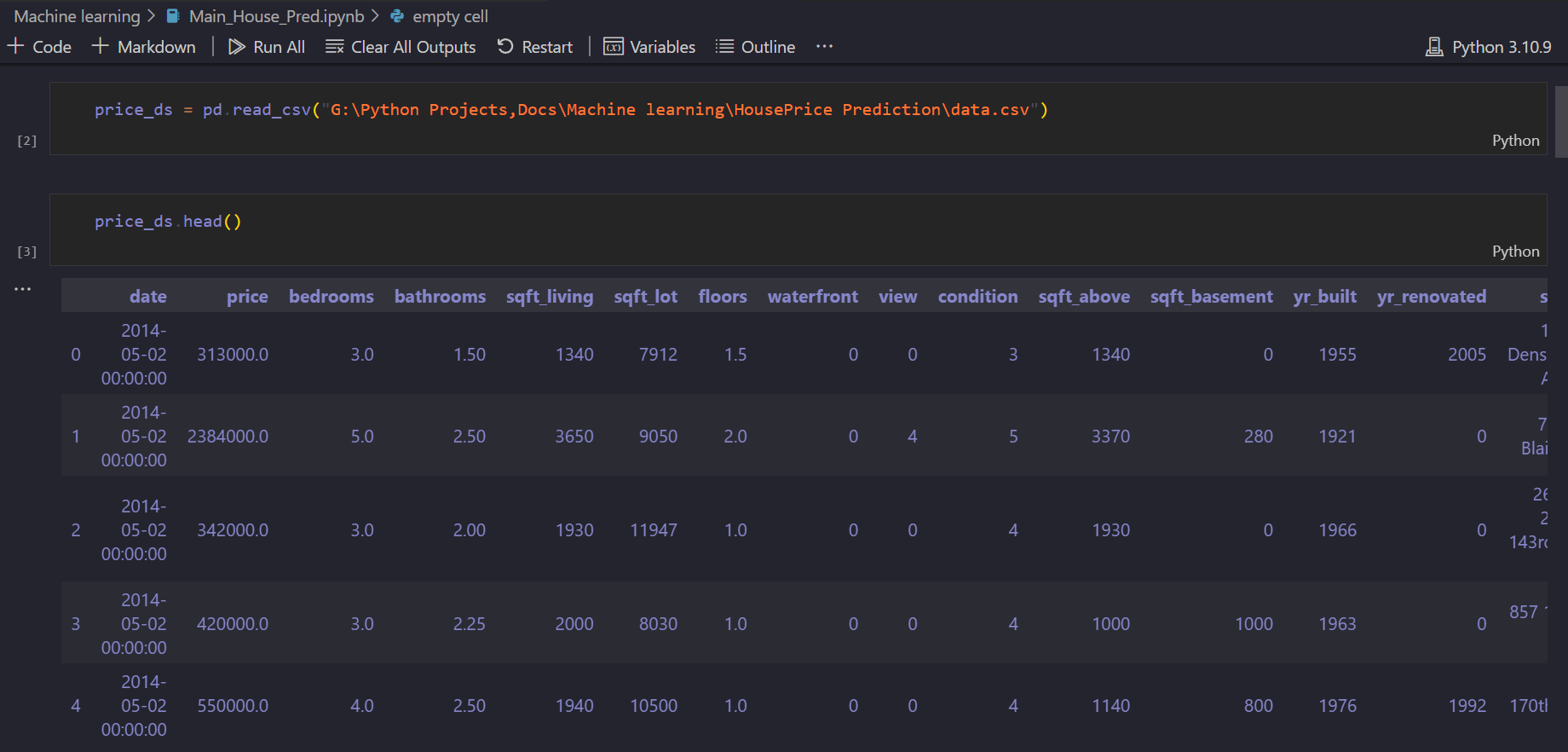
using a dataset. They specify the allowed uses of the data, attribution requirements, restrictions on redistribution or modification, access

limitations, ownership information, and any disclaimers or liabilities associated with the data. Understanding the data license helps ensure legal and ethical use of the dataset and guides your actions regarding sharing, analysis, and modifications.

**Found Datasets and their Sources**

|  |
| --- |
|  |

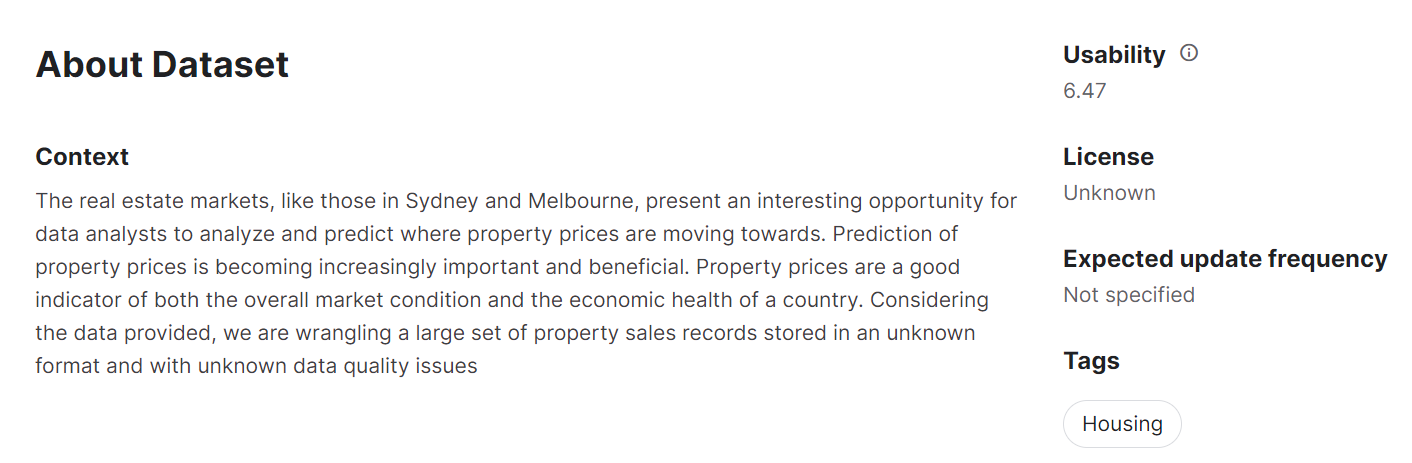
1. Sayan Mukherjee



**Source:**

**License** : unknown

**Data description** :



2. Ritaban Guchait

Dataset info:



Data Source: Amazon Car Accessories Dataset is a comprehensive collection of data related to car accessories available on the Amazon platform. This dataset encompasses a wide range of product information, including titles, prices, reviews, shipment details, stock updates, and offers. It serves as a valuable resource for researchers, analysts, and enthusiasts interested in studying consumer behavior, market trends, and product performance within the car accessories industry.

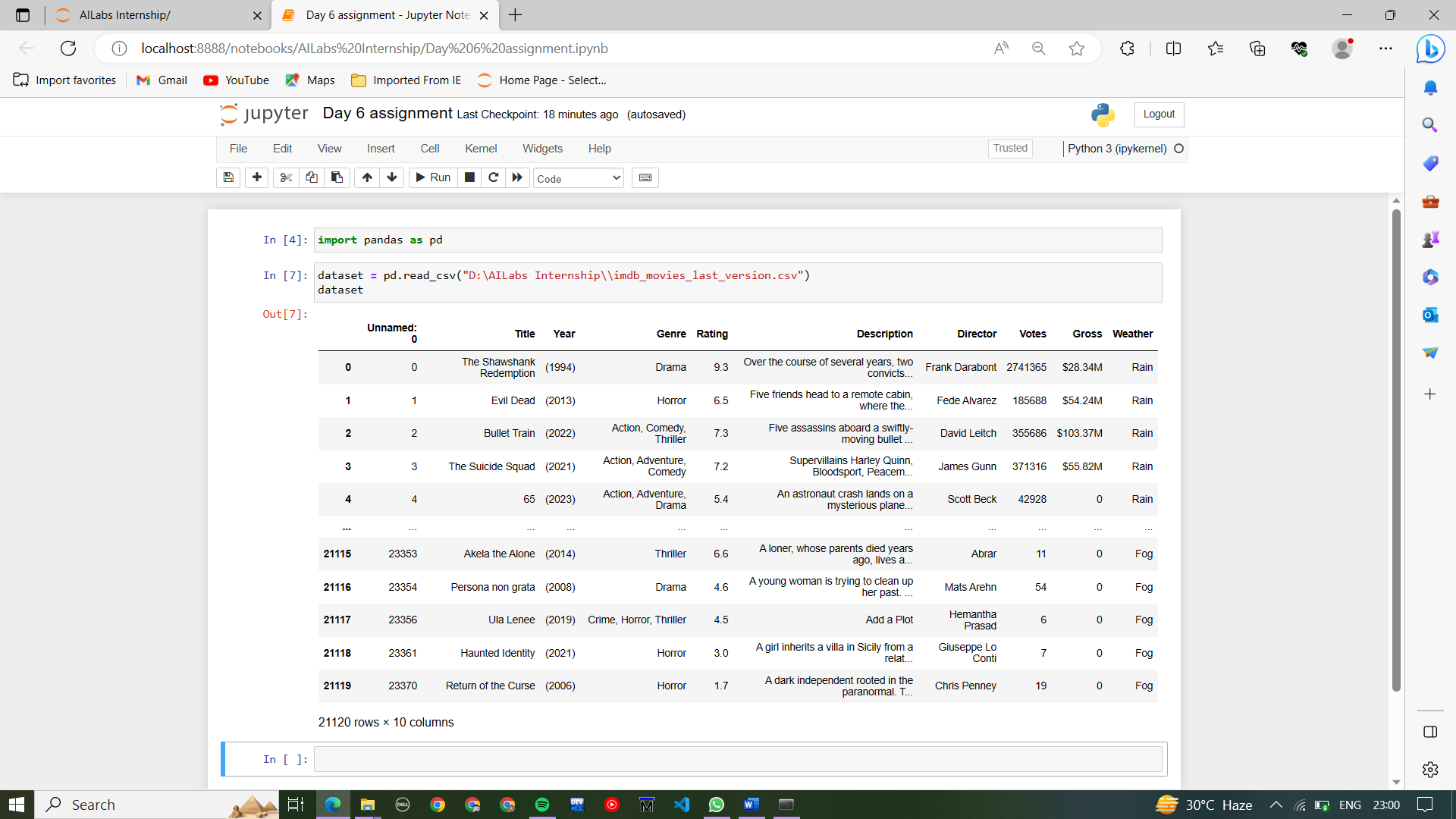
The dataset provides detailed information on various aspects of car accessories, enabling users to explore and analyze key variables. The "Title" column contains the names or descriptions of the car accessories available on Amazon. These titles offer insights into the types of products, brands, and features offered in the market.

Link of Datset: https://www.kaggle.com/datasets/rayhan32/amazon-car-accessories-dataset

Link of License: https://creativecommons.org/publicdomain/zero/1.0/

3. Anwesha Patra

Dataset info:



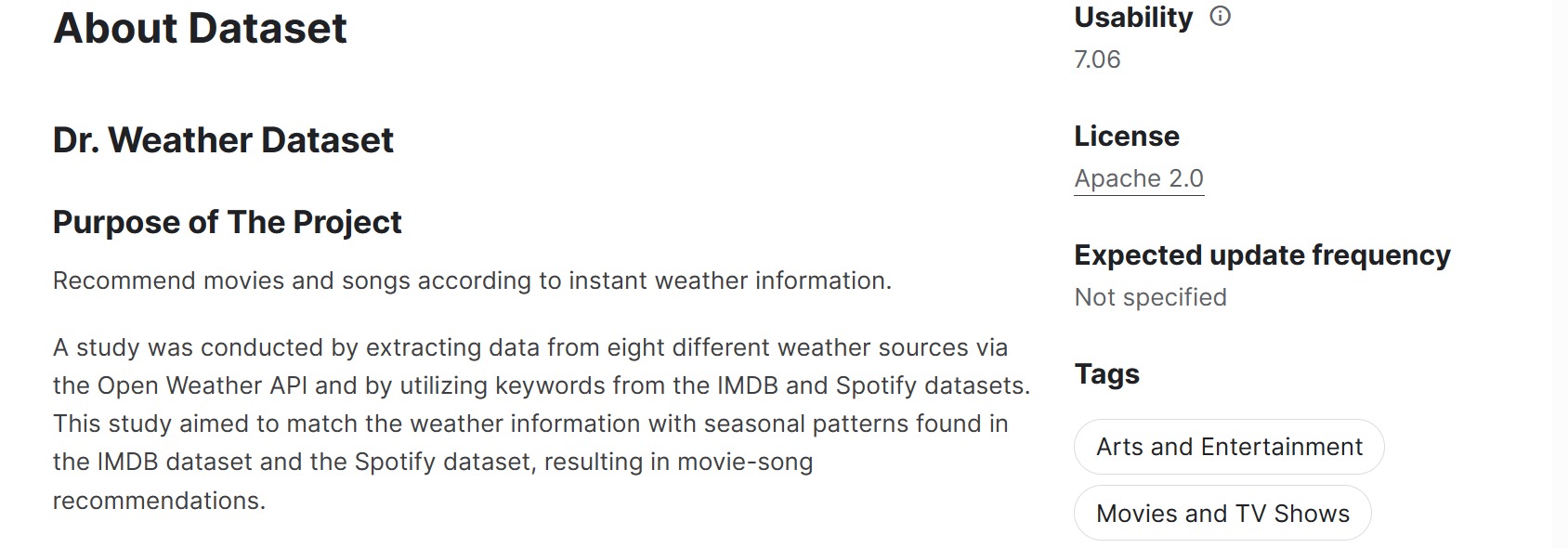
Data Source:

https://www.kaggle.com/datasets/receplyasolu/20k-weather-labeled-imdb-movies

License:

https://www.apache.org/licenses/LICENSE-2.0

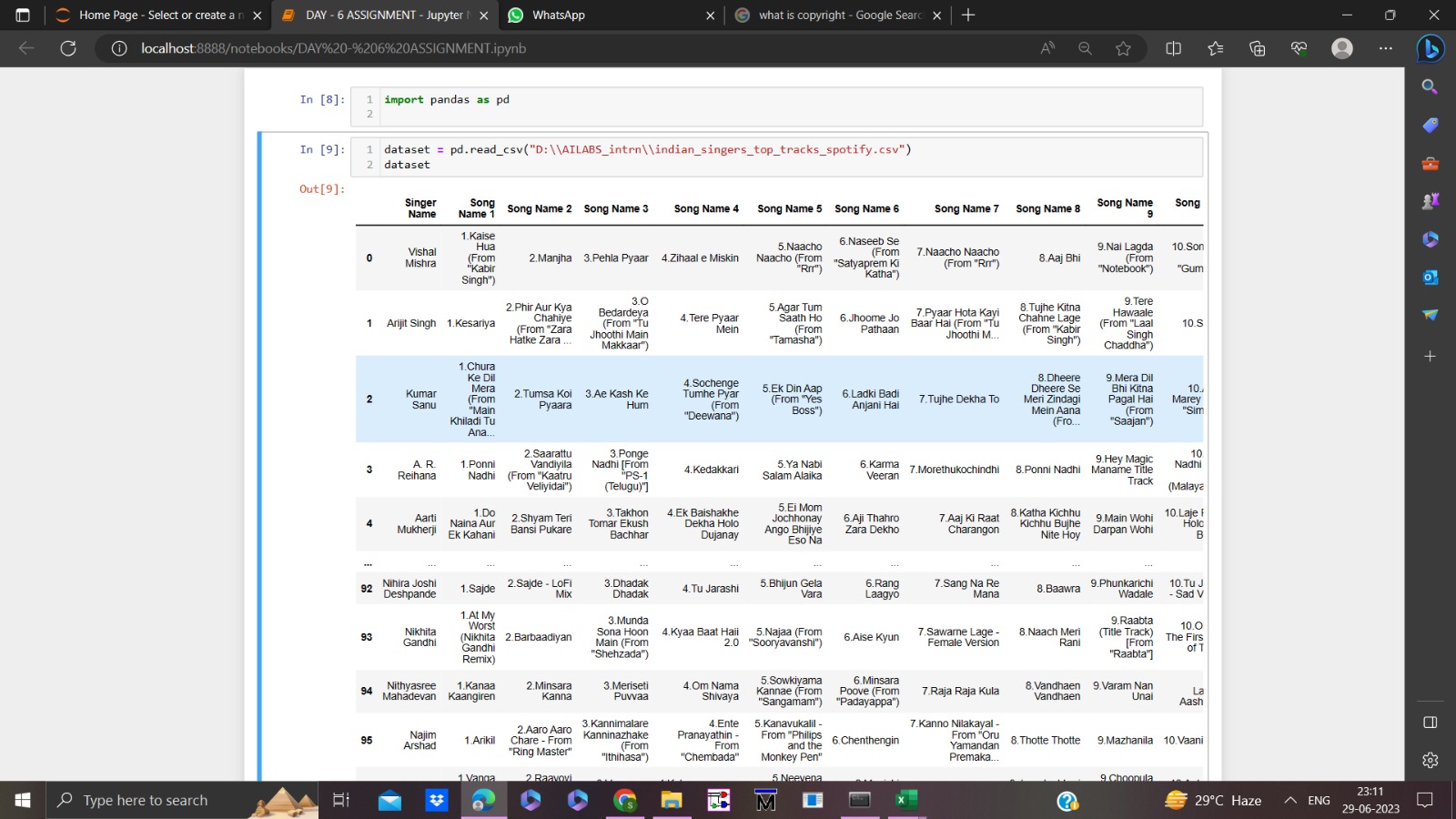
**Data description** :



Source Description :

The **Apache License** is a [permissive](https://en.wikipedia.org/wiki/Permissive_free_software_licence) [free software license](https://en.wikipedia.org/wiki/Free-software_license) written by the [Apache Software Foundation](https://en.wikipedia.org/wiki/Apache_Software_Foundation) (ASF). It allows users to use the software for any purpose, to distribute it, to modify it, and to distribute modified versions of the software under the terms of the license, without concern for [royalties](https://en.wikipedia.org/wiki/Royalties). The ASF and its projects release their software products under the Apache License. The license is also used by many non-ASF projects.4. Swagata Nanda

Dataset info:



Data Source: https://www.kaggle.com/datasets/mauryansshivam/spotify-top-10-tracks-singer-wise-in-india

License : https://creativecommons.org/publicdomain/zero/1.0/

Description:

About Dataset:

Spotify Top 10 Tracks Singer Wise in India contains Top 10 Songs of 97 Singers (mostly Indian singers) in Indian Market of Spotify. These 97 singers has song in different languages of India as well as English also at Spotify.

**This data is in tabular form, but it contains mainly textual data.**

## Usability

9.41

License

[CC0: Public Domain](https://creativecommons.org/publicdomain/zero/1.0/)

Expected update frequency

Annually

Tags

[Music](https://www.kaggle.com/datasets?tags=2304-Music), [Tabular](https://www.kaggle.com/datasets?tags=14101-Tabular), [Beginner](https://www.kaggle.com/datasets?tags=13102-Beginner), [India](https://www.kaggle.com/datasets?tags=3023-India), [English](https://www.kaggle.com/datasets?tags=16895-English)

Source Description :

Dataset can be two type :

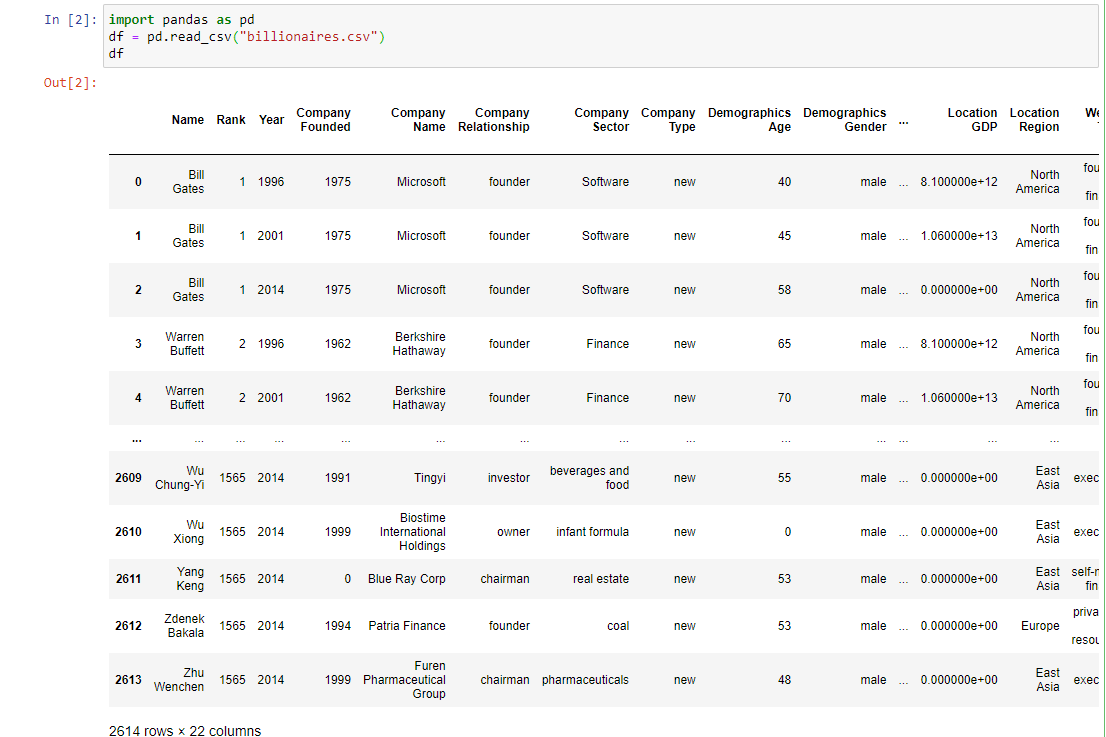
1. open source

2. Paid

This spotify data set is open source . Open-source software (OSS) is non-proprietary software that allows anyone to modify, distribute, enhance, or simply view its source code. Open source is simply a license to use and modify the software under the terms of the open source software agreement, such as a General Public License (GPL).

5.Soumik Mukherjee

Dataset info:



Data Source:[World Top Billionaires | Kaggle](https://www.kaggle.com/datasets/bhanupratapbiswas/world-top-billionaires?resource=download)

About Dataset:

As of my knowledge cutoff in September 2021, here are some of the world's top billionaires. Please note that this information may be outdated, and rankings can change over time due to fluctuations in net worth.

1. Jeff Bezos - Founder of Amazon.com

2. Elon Musk - CEO of Tesla and SpaceX

3. Bernard Arnault - CEO of LVMH (Louis Vuitton Moët Hennessy)

4. Bill Gates - Co-founder of Microsoft

5. Mark Zuckerberg - Co-founder and CEO of Facebook

6. Warren Buffett - Chairman and CEO of Berkshire Hathaway

7. Larry Ellison - Co-founder of Oracle Corporation

8. Larry Page - Co-founder of Google (Alphabet Inc.)

9. Sergey Brin - Co-founder of Google (Alphabet Inc.)

10. Mukesh Ambani - Chairman and largest shareholder of Reliance Industries Limited

It's important to keep in mind that these rankings may have changed since 2021, and new billionaires may have entered the list. I recommend referring to the latest Forbes billionaire list or other reliable sources for the most up-to-date information on the world's top billionaires.

Researchers have compiled a multi-decade database of the super-rich. Building off the Forbes World’s Billionaires lists from 1996-2014, scholars at Peterson Institute for International Economics have added a couple dozen more variables about each billionaire - including whether they were self-made or inherited their wealth. (Roughly half of the European billionaires and one-third of U.S. billionaires got a significant financial boost from family, the authors estimate.)

**Usability**

info

10.00

## License

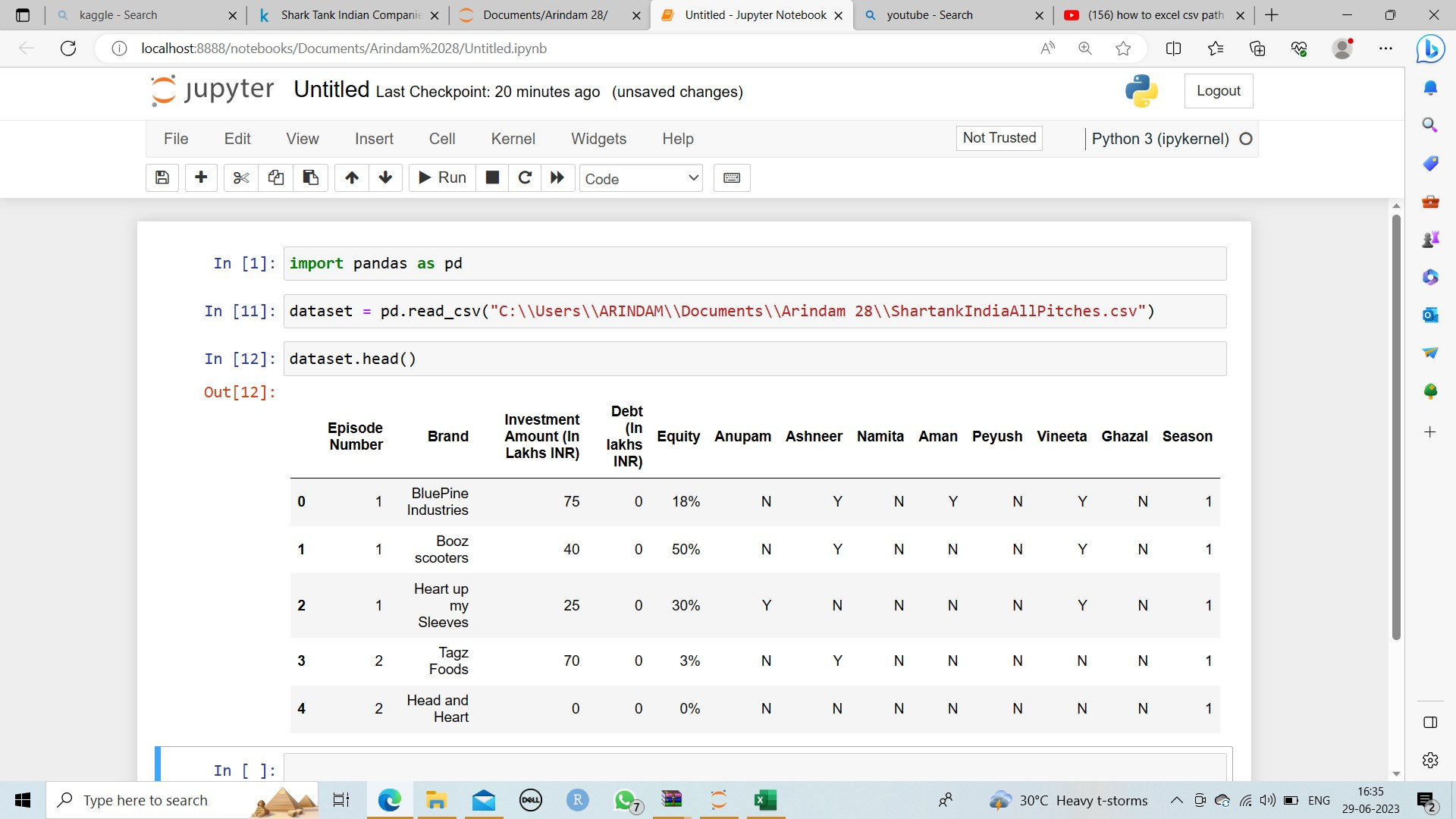
[Apache 2.0](https://www.apache.org/licenses/LICENSE-2.0)

## Expected update frequency

Never

6. Arindam Pakhira

Dataset info:



Data Source:

Shark Tank recently came to India and became a massively popular show replacing the stereotypical family drama.  
So I wanted to make a Dataset that keeps to date with all the Pitches that happen on Shark Tank India

There are 14 columns with Episode Number, Pitch Id, Name of the Brand, Brief Idea of the Brand, The amount invested by the sharks, The amount of debt given (if any), the amount of equity diluted by the brand, and the Next seven columns correspond to the Sharks and the Y states if they have invested in that Brand or not. The last column is for the Season Number

# License : https://creativecommons.org/publicdomain/zero/1.0/

## Usability

10.00

License

[CC0: Public Domain](https://creativecommons.org/publicdomain/zero/1.0/)

Expected update frequency

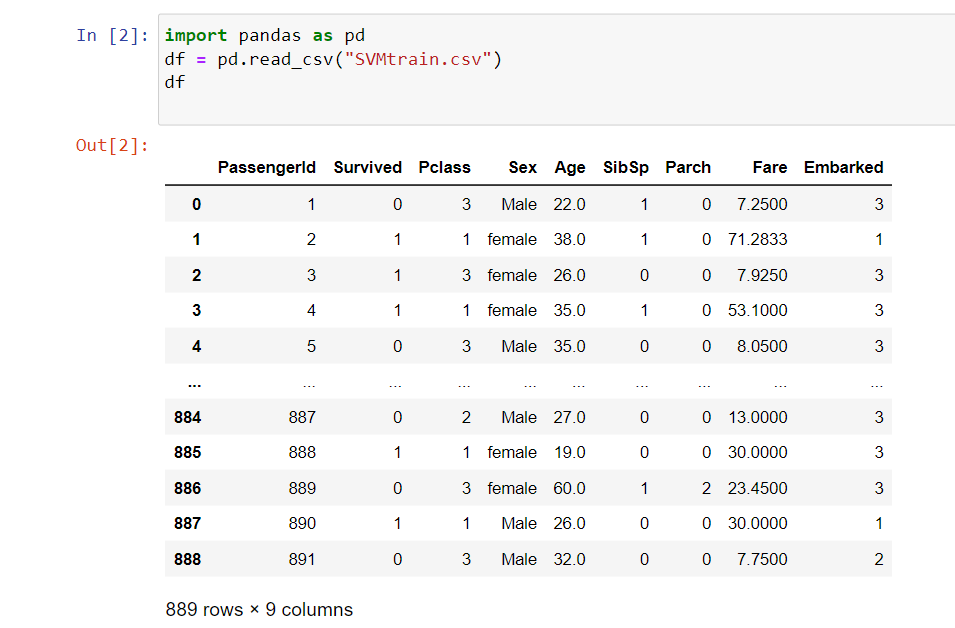
Quarterly

## Tags

[Business](https://www.kaggle.com/datasets?tags=11102-Business) [Arts and Entertainment](https://www.kaggle.com/datasets?tags=2200-Arts+and+Entertainment) [Tabular](https://www.kaggle.com/datasets?tags=14101-Tabular) [Beginner](https://www.kaggle.com/datasets?tags=13102-Beginner) [India](https://www.kaggle.com/datasets?tags=3023-India)

7.Aritra Mukherjee

Dataset info:



Data Source:[Titanic Dataset | Kaggle](https://www.kaggle.com/datasets/shubhamgupta012/titanic-dataset)

About Dataset:

The Titanic Passenger dataset provides information about passengers who were aboard the RMS Titanic during its ill-fated maiden voyage. This dataset is often used for exploring patterns and factors associated with survival on the Titanic.

The dataset includes the following columns:

PassengerId: Unique identifier for each passenger.  
Survived: Survival status of the passenger (0 = Not Survived, 1 = Survived).  
Pclass: Passenger class (1 = First class, 2 = Second class, 3 = Third class).  
Sex: Gender of the passenger.  
Age: Age of the passenger.  
SibSp: Number of siblings/spouses aboard the Titanic.  
Parch: Number of parents/children aboard the Titanic.  
Fare: Fare paid by the passenger.  
Embarked: Port of embarkation (C = Cherbourg, Q = Queenstown, S = Southampton).

## Usability

info

This score is calculated by Kaggle.

10.00

## License

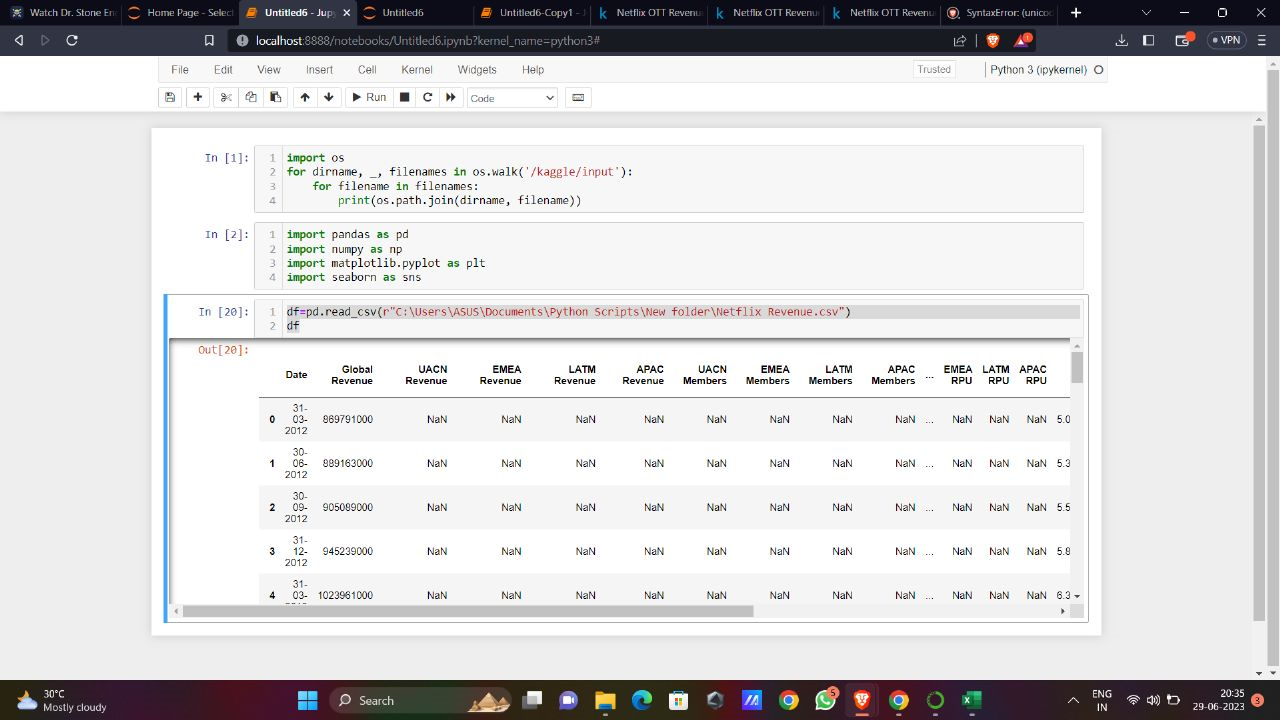
Other (specified in description)

## Expected update frequency

Never

8. Ayan Kundu

Dataset info:



|  |
| --- |
| License:  Creative commons corporation is not a law firm and does not provide legal services. Distribution of this document does not create an attorney-client relationship. Creative common provides this information on “as-is” basis. Creative common makes no warranties  Regarding the use of this document or the information or works provide hereunder, and disclaims liability for damages resulting from the use of this document or the information or works provides hereunder |

|  |
| --- |
| Data description:  About dataset:  Netflix, Incorporated, a business based in Los Gatos, California,  owns and runs the American subscription video on-demand  over-the-top streaming television service. It provides a wide  range of genre-specific films and television shows, and it  supports numerous languages .Netflix was introduced on January 16,2007, over a decade after Netflix, Inc. began its DVD-by-mail  mail service. It is the most popular video on demand streaming service, with over 232 million paid customers across more than 190  countries. |

|  |
| --- |
| Link:  <https://creativecommons.org/publicdomain/zero/1.0/legalcode> |

9.Biswanath Mandal

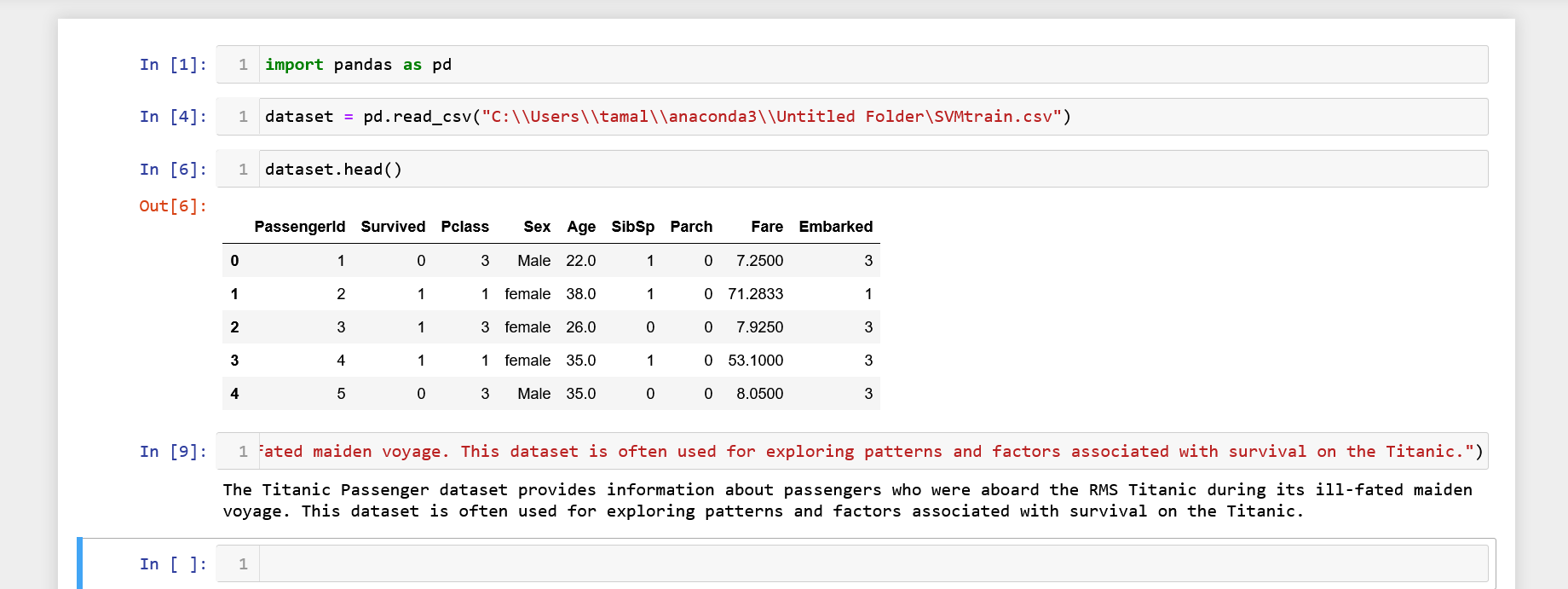
Dataset info:



Data Source:

10.Tamal Mondal

Dataset info:



Data Source:

License : None

Data description :

The Titanic Passenger dataset provides information about passengers who were aboard the RMS Titanic during its ill-fated maiden voyage. This dataset is often used for exploring patterns and factors associated with survival on the Titanic.