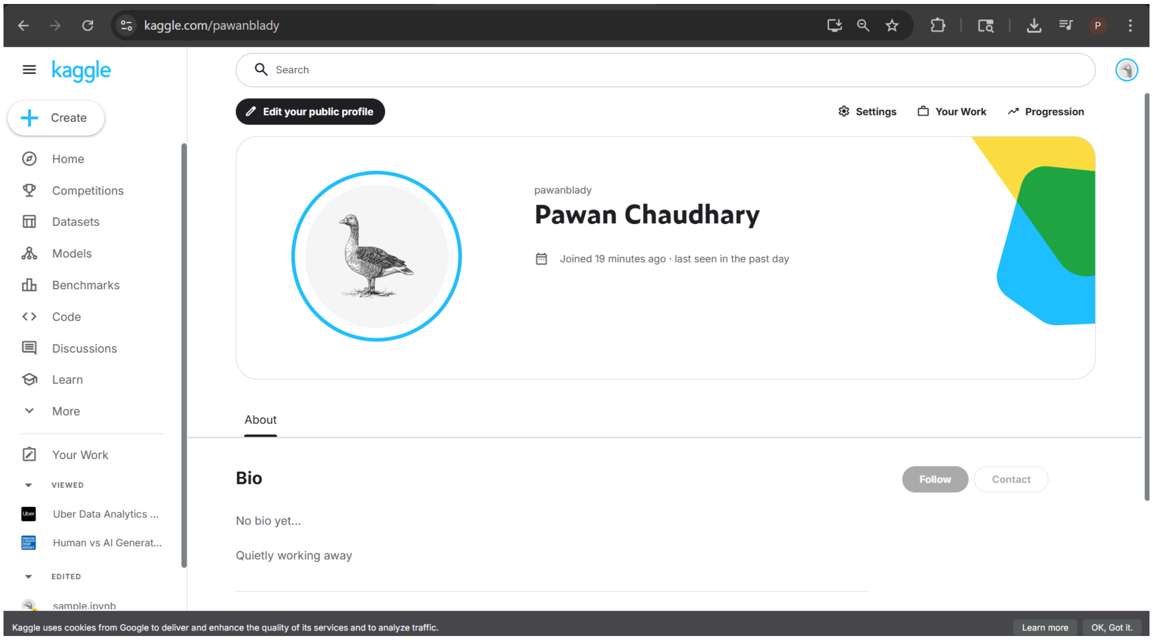
Pawan Chaudhary

1. Exploring Kaggle

In [5]:
```
# Kaggle Profile
```



pawanblady

Pawan Chaudhary

Joined 19 minutes ago · last seen in the past day

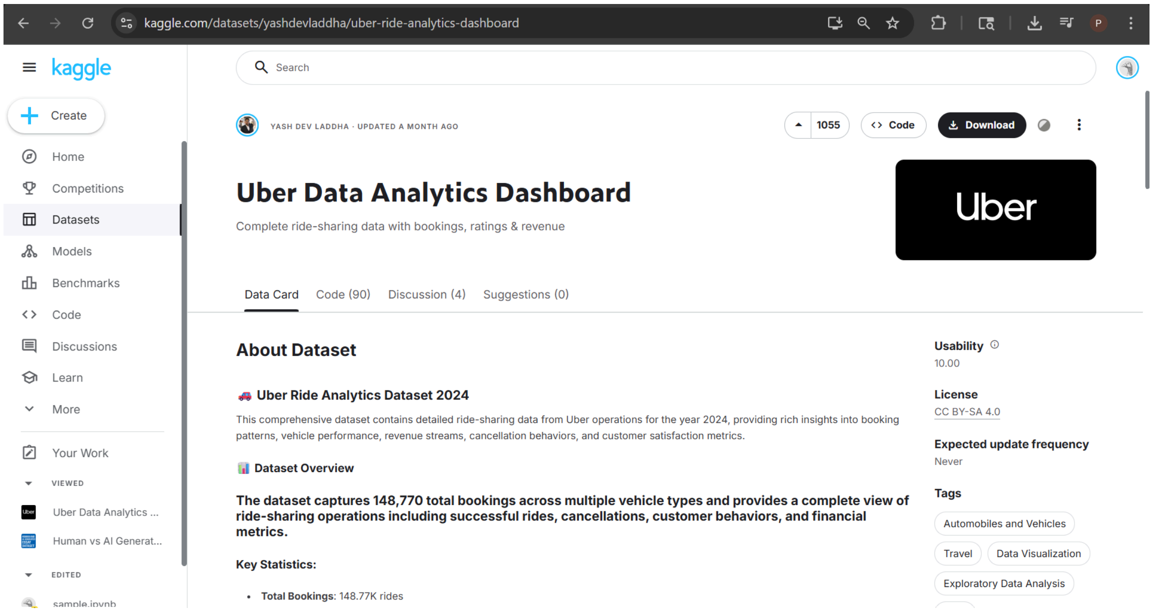
Bio

No bio yet...

Quietly working away

Kaggle uses cookies from Google to deliver and enhance the quality of its services and to analyze traffic.

In [3]:
```
# Most Popular Dataset
```



Uber Data Analytics Dashboard

Complete ride-sharing data with bookings, ratings & revenue

About Dataset

🚗 Uber Ride Analytics Dataset 2024

This comprehensive dataset contains detailed ride-sharing data from Uber operations for the year 2024, providing rich insights into booking patterns, vehicle performance, revenue streams, cancellation behaviors, and customer satisfaction metrics.

📊 Dataset Overview

The dataset captures 148,770 total bookings across multiple vehicle types and provides a complete view of ride-sharing operations including successful rides, cancellations, customer behaviors, and financial metrics.

Key Statistics:

- **Total Bookings:** 148.77K rides

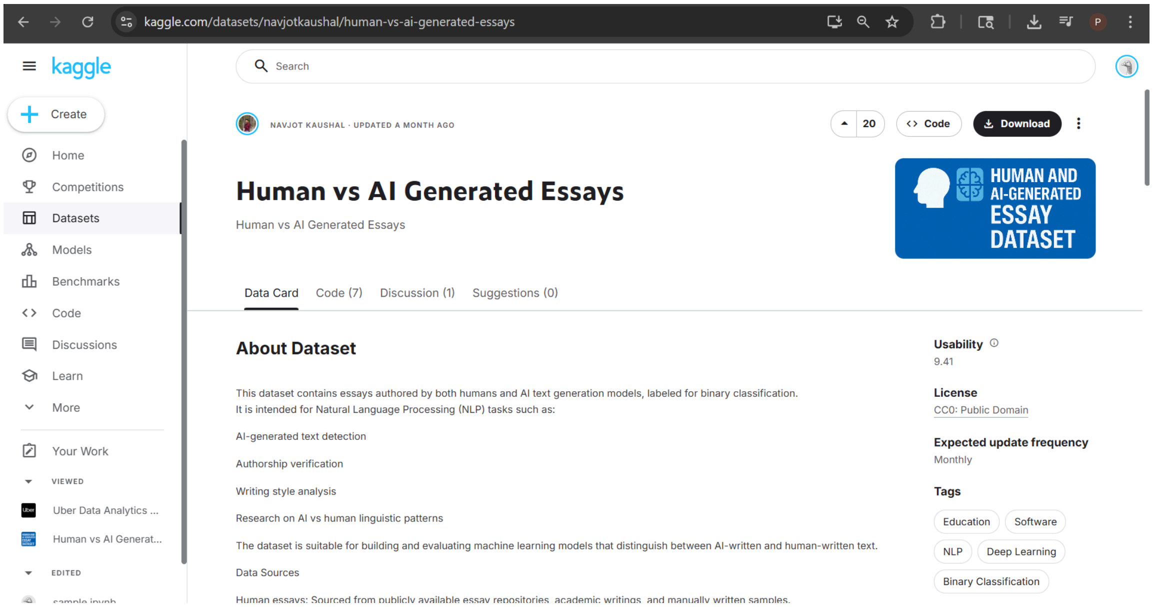
Usability
10.00

License
CC BY-SA 4.0

Expected update frequency
Never

Tags
Automobiles and Vehicles
Travel
Data Visualization
Exploratory Data Analysis

In [4]:
```
# Highest Trending Dataset
```



Human vs AI Generated Essays

Human vs AI Generated Essays

About Dataset

This dataset contains essays authored by both humans and AI text generation models, labeled for binary classification.
It is intended for Natural Language Processing (NLP) tasks such as:

AI-generated text detection

Authorship verification

Writing style analysis

Research on AI vs human linguistic patterns

The dataset is suitable for building and evaluating machine learning models that distinguish between AI-written and human-written text.

Data Sources

Human essays: Sourced from publicly available essay repositories, academic writings, and manually written samples.

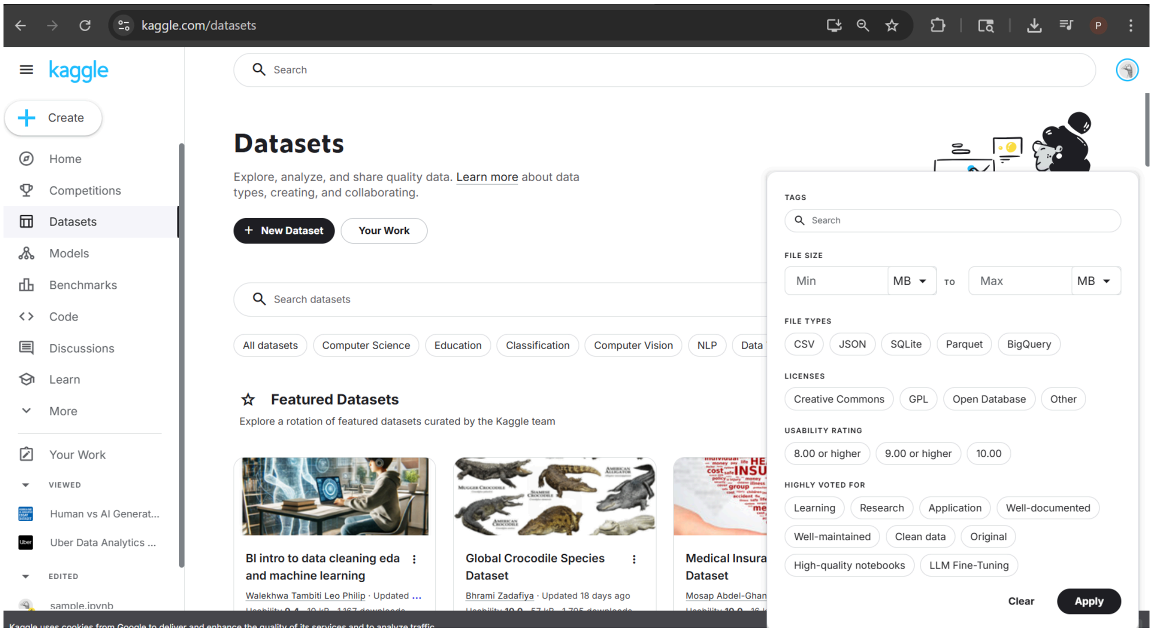
Usability
9.41

License
CC0: Public Domain

Expected update frequency
Monthly

Tags
Education Software NLP Deep Learning Binary Classification

In [ ]:
```
# Filter Option
```



Datasets

Explore, analyze, and share quality data. Learn more about data types, creating, and collaborating.

Featured Datasets

Explore a rotation of featured datasets curated by the Kaggle team

BI intro to data cleaning eda and machine learning
Walekhwa Tambiti Leo Philip · Updated ...

Global Crocodile Species Dataset
Bhrami Zadafiya · Updated 18 days ago

Medical Insura Dataset
Mosap Abdel-Ghan

Kaggle uses cookies from Google to deliver and enhance the quality of its services and to analyze traffic.

TAGS
Search

FILE SIZE
Min MB to Max MB

FILE TYPES
CSV JSON SQLite Parquet BigQuery

LICENSES
Creative Commons GPL Open Database Other

USABILITY RATING
8.00 or higher 9.00 or higher 10.00

HIGHLY VOTED FOR
Learning Research Application Well-documented Well-maintained Clean data Original High-quality notebooks LLM Fine-Tuning

Clear Apply

2. Loading Data

In [ ]:
```
# Databases (SQLite)
```

I chose "Social Media Advertisement Performance" dataset, which analyzes a Simulated Ad Funnel with User & Campaign Data

In [2]:
```
import pandas as pd
```

In [3]:
```
import sqlalchemy as sqla
```

```
In [4]: db_path = '/kaggle/input/social-media-advertisement-performance/ad_campaign_db.sqli
```

```
In [5]: db = sqla.create_engine(f'sqlite:///{db_path}')
```

```
In [6]: query_for_tables = "SELECT name FROM sqlite_master WHERE type='table';"
```

```
In [8]: table_names = pd.read_sql(query_for_tables, db)
```

```
In [9]: print("Tables in the database:")
        print(table_names)
```

```
Tables in the database:
        name
0      users
1  campaigns
2        ads
3  ad_events
```

```
In [10]: query_sqlite = "SELECT * FROM campaigns LIMIT 5"
         df = pd.read_sql(query_sqlite, db)
         print(df)
```

```
   campaign_id               name  start_date    end_date  duration_days  \
0            1  Campaign_1_Launch  2025-05-25  2025-07-23             59   
1            2  Campaign_2_Launch  2025-04-16  2025-07-07             82   
2            3  Campaign_3_Winter  2025-05-04  2025-06-29             56   
3            4  Campaign_4_Summer  2025-06-04  2025-08-08             65   
4            5  Campaign_5_Launch  2025-07-11  2025-08-28             48   

   total_budget  
0      24021.32  
1      79342.41  
2      14343.25  
3      45326.60  
4      68376.69  
```

```
In [11]: # Flat Files (CSV)
```

```
In [12]: import pandas as pd
```

```
In [20]: pd.options.display.max_columns = None
         pd.options.display.width = None
         pd.options.display.float_format = '{:,.2f}'.format
```

```
In [21]: file_path = '/kaggle/input/uber-ride-analytics-dashboard/ncr_ride_bookings.csv'
```

```
In [22]: df = pd.read_csv(file_path)
```

```
In [23]: first_five_rows = df[0:5]
         print(first_five_rows)
         Date      Time    Booking ID   Booking Status   Customer ID  \
0  2024-03-23  12:29:38  "CNR5884300"  No Driver Found  "CID1982111"   
1  2024-11-29  18:01:39  "CNR1326809"       Incomplete  "CID4604802"   
2  2024-08-23  08:56:10  "CNR8494506"        Completed  "CID9202816"   
3  2024-10-21  17:17:25  "CNR8906825"        Completed  "CID2610914"   
4  2024-09-16  22:08:00  "CNR1950162"        Completed  "CID9933542"   

    Vehicle Type      Pickup Location      Drop Location  Avg VTAT  Avg CTAT  \
0          eBike          Palam Vihar            Jhilmil       NaN       NaN   
1       Go Sedan        Shastri Nagar  Gurgaon Sector 56      4.90     14.00   
2           Auto              Khandsa      Malviya Nagar     13.40     25.80   
3  Premier Sedan  Central Secretariat           Inderlok     13.10     28.50   
4           Bike     Ghitorni Village        Khan Market      5.30     19.60   

   Cancelled Rides by Customer Reason for cancelling by Customer  \
0                          NaN                               NaN   
1                          NaN                               NaN   
2                          NaN                               NaN   
3                          NaN                               NaN   
4                          NaN                               NaN   

   Cancelled Rides by Driver Driver Cancellation Reason  Incomplete Rides  \
0                        NaN                        NaN               NaN   
1                        NaN                        NaN              1.00   
2                        NaN                        NaN               NaN   
3                        NaN                        NaN               NaN   
4                        NaN                        NaN               NaN   

  Incomplete Rides Reason  Booking Value  Ride Distance  Driver Ratings  \
0                     NaN            NaN            NaN             NaN   
1       Vehicle Breakdown         237.00           5.73             NaN   
2                     NaN         627.00          13.58            4.90   
3                     NaN         416.00          34.02            4.60   
4                     NaN         737.00          48.21            4.10   

   Customer Rating Payment Method  
0              NaN            NaN  
1              NaN            UPI  
2             4.90     Debit Card  
3             5.00            UPI  
4             4.30            UPI  
```

In [31]:

```
# APIs
```

In [25]:

```
import requests
```

In [26]:

```
base_url = 'https://pokeapi.co/api/v2/pokemon'
```

In [27]:

```
resp = requests.get(base_url)
resp
```

Out[27]:

```
<Response [200]>
```

In [28]:

```
data = resp.json()
data['results'][0]['name']
```

```
Out[28]:  'bulbasaur'
```

```
In [29]:  pokemon = pd.DataFrame(data['results'], columns=['name'])
          pokemon.head()
```

```
Out[29]:          name
          0    bulbasaur
          1      ivysaur
          2     venusaur
          3   charmander
          4   charmeleon
```

An API is a tool that allows us to access data directly from a source in a quick and organized way. It is important for data analysis because it saves time, reduces errors, and ensures that we are working with accurate and up-to-date information.