

Machine Learning

Video 16:

Working with JSON/SQL:

JSON (JavaScript Object Notation) is a lightweight data-interchange format that is easy for humans to read and write. It uses key-value pairs to structure data, often represented as objects or arrays. JSON is commonly used for transmitting data between a server and web application.

Code: <https://github.com/campusx-official/100-days-of-machine-learning/blob/main/day16%20-%20working-with-json-and-sql/train.json>

Data: <https://github.com/campusx-official/100-days-of-machine-learning/tree/main/day16%20-%20working-with-json-and-sql>

<https://www.kaggle.com/datasets/busieltmorley/worldcities-pop-lang-rank-sql-create-tbls>

Video 17:

Fetching Data From an API:

An API (Application Programming Interface) is a set of rules that allows different software applications to communicate with each other. It defines the methods and data formats that applications can use to request and exchange information. APIs enable integration between different systems and services, making them essential for modern software development.

How a data looks when fetched from any API:

```
Pretty print ☒
{
  "page": 1,
  "results": [
    {
      "adult": false,
      "backdrop_path": "/zfbjgQE1uSd9wiPTX4VzsLi0rGG.jpg",
      "genre_ids": [18, 80],
      "id": 278,
      "original_language": "en",
      "original_title": "The Shawshank Redemption",
      "overview": "Imprisoned in the 1940s for the double murder of his wife and her lover, upstanding banker Andy Dufresne begins a new life at the Shawshank prison, where he puts his accounting skills to work for an amoral warden. During his long stretch in prison, Dufresne comes to be admired by the other inmates -- including an older prisoner named Red -- for his integrity and unquenchable sense of hope.",
      "popularity": 162.849,
      "poster_path": "/9cqNxx0GxF0bflZmeSMuL5tnGzr.jpg",
      "release_date": "1994-09-23",
      "title": "The Shawshank Redemption",
      "video": false,
      "vote_average": 8.7,
      "vote_count": 27621
    },
    {
      "adult": false,
      "backdrop_path": "/tmU7GeKVybMWFButWEGl2M4GeiP.jpg",
      "genre_ids": [18, 80],
      "id": 238,
      "original_language": "en",
      "original_title": "The Godfather",
      "overview": "Spanning the years 1945 to 1955, a chronicle of the fictional Italian-American Corleone crime family. When organized crime family patriarch, Vito Corleone barely survives an attempt on his life, his youngest son, Michael steps in to take care of the would-be killers, launching a campaign of bloody revenge.",
      "popularity": 161.779,
      "poster_path": "/3bhkrj58Vtu7enYsRolD1fZdja1.jpg",
      "release_date": "1972-03-14",
      "title": "The Godfather",
      "video": false,
      "vote_average": 8.689,
      "vote_count": 20949
    },
    {
      "adult": false,

```

To view it in readable form we will use the online websites:

```
the limitations on human space travel and conquer the vast distances involved in an interstellar voyage.",
  "popularity": 286.576,
  "poster_path": "/gEU2QniE6E77NI61CU6Mx1NBvIx.jpg",
  "release_date": "2014-11-05",
  "title": "Interstellar",
  "video": false,
  "vote_average": 8.4,
  "vote_count": 36319
},
{
  "adult": false,
  "backdrop_path": "/gavyCu1UaTaTNPsVaGXT6pe5u24.jpg",
  "genre_ids": [35, 18],
  "id": 637,
  "original_language": "it",
  "original_title": "La vita è bella",
  "overview": "A touching story of an Italian book seller of Jewish ancestry who lives in his own little fairy tale. His creative and happy life would come to an abrupt halt when his entire family is deported to a concentration camp during World War II. While locked up he tries to convince his son that the whole thing is just a game.",
  "popularity": 52.356,
  "poster_path": "/74hLDKjD5aGY0ot06esUVaeISa2.jpg",
  "release_date": "1997-12-20",
  "title": "Life Is Beautiful",
  "video": false,
  "vote_average": 8.448,
  "vote_count": 13155
}
],
"total_pages": 497,
"total_results": 9929
}
```

Name	Value
page	1
results	...
total_pages	497
total_results	9929

Code: <https://github.com/campusx-official/100-days-of-machine-learning/blob/main/day16%20-%20working-with-json-and-sql/train.json>

Video 18:

Fetching data using Web Scraping:

NA

Video 19:

Understanding Your Data:

Asking basic questions? How to analyse the data?

1. Importing Data with Pandas:

- Use `pd.read_csv('file_path')` to read CSV data into a DataFrame.
- Example: `df = pd.read_csv('train.csv')`

2. Check the Size of the Data:

- Use `df.shape` to get the number of rows and columns.
- Example output: `(891, 12)` indicates 891 rows and 12 columns.

3. Preview the Data:

- Use `df.head()` to see the first 5 rows of the dataset.
- Use `df.sample(n)` to get n random rows from the dataset.
- Example: `df.sample(5)` will return 5 random rows.

4. Check Data Types of Columns:

- Use `df.info()` to check data types and non-null counts for each column.
- Example output: It shows column names, data types (int64, float64, object), and null values.

5. Missing Data:

- Use `df.isnull().sum()` to check for missing values in each column.
- Example output: Age has 177 missing values, Cabin has 687, Embarked has 2.

6. Statistical Summary:

- Use `df.describe()` to get summary statistics like mean, standard deviation, min, max, etc., for numeric columns.

7. Check for Duplicates:

- Use `df.duplicated().sum()` to check for duplicate rows in the dataset.
- Example: If output is 0, no duplicates exist.

8. Correlation Between Columns:

- Use `df.corr(numeric_only=True)` to see the correlation between numeric columns.
- Use `df.corr(numeric_only=True)['ColumnName']` to see the correlation of a specific column (e.g., "Survived") with others.

Code link:

https://colab.research.google.com/drive/1uKX9vLGgkl0o2jmE_x6qCO_HSEcPt6W5?usp=sharing

Video 20:

EDA using Univariate Analysis:

Exploratory	Data	Analysis	(EDA):
EDA is a data analysis approach that uses statistical graphics, plots, and summary statistics to explore and understand the structure, patterns, and relationships within a dataset. It helps in identifying anomalies, trends, and data distributions, guiding further analysis and model-building decisions.			

Univariate	Analysis:
Univariate analysis focuses on the examination of a single variable. It involves analyzing the distribution, central tendency (mean, median, mode), dispersion (variance, standard deviation), and shape (skewness, kurtosis) to understand the data's characteristics and detect outliers or patterns in isolation from other variables.	

Code:

<https://colab.research.google.com/drive/1PdnCg5BkDZIT4fZffB7IcB-6FY3AgIfx?usp=sharing>