In Python create and use setup environment

Use the below commands to create and environment

* python -m venv myenv
* myenv\Scripts\activate

1. **ETL Scripts:** Python scripts for data extraction, transformation, and loading.

**Data Extraction**

* Install Kaggle API and link it with your account
* Use command: kaggle competitions download -c titanic

The above command is used to download the titanic dataset

* Use Command: unzip titanic.zip

This command is used to unzip the file downloaded

* Further extraction steps are available in the ipynb file attached for extraction

**Data Transformation and Data Loading**

* The python steps are provided in the ipynb file attached with description

1. **Database Schema**: Schema design for the chosen storage system.

**Passengers Table**:

* **passenger\_id**: INT, Primary Key
* **survived**: BOOLEAN
* **pclass**: INT
* **name**: VARCHAR(255)
* **sex**: VARCHAR(10)
* **age**: FLOAT
* **siblings\_spouses\_aboard**: INT
* **parents\_children\_aboard**: INT
* **fare**: FLOAT
* **embarked**: VARCHAR(10)

**Schema**

CREATE TABLE Passengers (

passenger\_id INT PRIMARY KEY,

survived BOOLEAN,

pclass INT,

name VARCHAR(255),

sex VARCHAR(10),

age FLOAT,

siblings\_spouses\_aboard INT,

parents\_children\_aboard INT,

fare FLOAT,

embarked VARCHAR(10)

);

1. **ML Pipeline Script:** Python script for the ML pipeline integration.

Python script is available in attached ipynb file with comments in each step

1. **Schema of the generated database**

|  |  |
| --- | --- |
| **Column Name** | **Data Type** |
| Age | FLOAT |
| Fare | FLOAT |
| FamilySize | FLOAT |
| Sex\_female | FLOAT |
| Sex\_male | FLOAT |
| Embarked\_C | FLOAT |
| Embarked\_Q | FLOAT |
| Embarked\_S | FLOAT |
| Title\_Capt | FLOAT |
| Title\_Col | FLOAT |
| Title\_Don | FLOAT |
| Title\_Dr | FLOAT |
| Title\_Jonkheer | FLOAT |
| Title\_Lady | FLOAT |
| Title\_Major | FLOAT |
| Title\_Master | FLOAT |
| Title\_Miss | FLOAT |
| Title\_Mlle | FLOAT |
| Title\_Mme | FLOAT |
| Title\_Mr | FLOAT |
| Title\_Mrs | FLOAT |
| Title\_Ms | FLOAT |
| Title\_Rev | FLOAT |
| Title\_Sir | FLOAT |
| Title\_the Countess | FLOAT |