

File Edit Selection View Go Run Terminal Help

Real Time Projects

Python 3.10.11

EXPLORER

REAL TIME PROJECTS

EDA_LLM- Automation Mistral

gradio

app.py

code.ipynb

data.csv

LIST OF LLM MODELS_1.xlsx

LIST OF LLM MODELS.xlsx

test.py

titanic_dataset_final.csv

Workshop Notes.txt

Workshop Notes.txt.bak

HEIGHT & WEIGHT

HOUSING REGRESSOR

templates

ANN.pkl

app.py

ElasticNet.pkl

house_area_column.txt

KNN.pkl

LassoRegression.pkl

LGBM.pkl

LinearRegression.pkl

model_evaluation_re...

model.py

PolynomialRegression.pkl

RandomForest.pkl

RidgeRegression.pkl

RobustRegression.pkl

SGDRegressor.pkl

SVM.pkl

USA_Housing.csv

XGBoost.pkl

ANN.pkl

ElasticNet.pkl

KNN.pkl

LassoRegression.pkl

LGBM.pkl

OUTLINE

TIMELINE

test.py

code.ipynb

EDA_LLM- Automation Mistral gradio

code.ipynb

import gradio as gr

def eda_analysis(file):

df = pd.read_csv(file.name)

summary = df.describe().to_string()

insights = generate_insights(summary)

return insights

Create Web Interface

demo = gr.Interface(fn=eda_analysis, inputs="file", outputs="text", title="AI-Powered EDA with Mistral")

Launch App

demo.launch()

Use share=True for Google Colab

Chat (CTRL + I) / Share (CTRL + L)

[*] ✓

* Running on local URL: <http://127.0.0.1:7863>

* To create a public link, set `share=True` in `launch()`.

AI-Powered EDA with Mistral

file

clean_data.csv

262.0 B

Clear

Submit

output

The dataset summary provides a statistical overview of the distribution of four variables: Unnamed: 0, Age, Salary, and Exp. Here's an interpretation of each variable based on the provided summary statistics:

1. **Unnamed: 0**: This column appears to be an index column with no distinct pattern or relationship with other columns in this dataset.

2. **Age**: The mean age is 2.5 years, which could indicate that the data represents employees at various stages of their careers rather than individuals. The minimum age is 0, but since it's a salary dataset, it might be an error or possibly representing interns or fresh graduates. The maximum age is 5 years, suggesting the data was collected over a short period.

3. **Salary**: The mean salary is around \$50,167 per year. The median salary (\$50,000) is quite close to the mean, indicating a relatively symmetrical distribution of salaries. However, the standard deviation (\$10,907) suggests some variation in salaries. The minimum and maximum salaries are \$34,000 and \$67,000 respectively, indicating wide disparities among employees' earnings.

4. **Exp**: The mean experience is approximately 4.7 years, suggesting a relatively young

master*

6 16

BLACKBOX Chat

Add Logs

CyberCoder

Improve Code

Share Code Link

Open Website

Spaces: 4

CRLF

Cell 7 of 7

AI Code Chat

25°C

Light rain

Search

26-07-2025

15:22

Home

EDA_LLM-Automation

AI-Powered EDA with Mistral

LLM-Powered Exploratory Data Analysis (EDA)

127.0.0.1:7860

Sign in

LLM-Powered Exploratory Data Analysis (EDA)

Upload any dataset CSV file and get automated EDA insights with AI-powered analysis and visualizations.

file_path

titanic_dataset_final.csv59.8 KB

Clear

Submit

EDA Report

4. ****Mean, Std, Min, 25th Percentile, Median (50th percentile), 75th Percentile, Max****: These values provide statistical insights about the data distribution for various continuous variables like Age, Fare, and some discrete ones like Pclass and Cabin (although Cabin seems to have missing values). For example:

- The mean age of passengers is approximately 29.36 years.
- More than half of the passengers were in 1st, 2nd, or 3rd class (Pclass=1, 2, or 3), as suggested by the median value of 3 for the Pclass column.
- The majority of passengers had an age between 22 and 35 years ($22 < \text{Age} < 35$).
- The minimum age is 0.42 years, which could be a data entry error or represent an infant.
- The maximum fare is 512.33 pounds, which was likely paid by a wealthy passenger.

5. ****Top Values****: The top value for the Name column is "Dooley, Mr. Patrick", indicating that there's one such person in this dataset. Similarly, the ticket number 347082 appears most frequently (7 times). However, keep in mind that these insights may be affected by duplicate entries in some columns like Ticket and Name.

6. ****Embarked****: The Embarked column shows the port from which passengers embarked on their Titanic voyage. The three possible values are C (Cherbourg), Q (Queenstown), or S (Southampton). The distribution suggests that most passengers embarked from Southampton, while a smaller percentage embarked from Cherbourg and Queenstown.

Data Visualizations

Distribution of PassengerId

Distribution of Survived

25°C Rain

Search

ENG IN

15:44 26-07-2025

Home

EDA_LLM-Automation

AI-Powered EDA with Mistral

LLM-Powered Exploratory Data Analysis

127.0.0.1:7860

Sign in

frequently (7 times). However, keep in mind that these insights may be affected by duplicate entries in some columns like Ticket and Name.

6. ****Embarked****: The Embarked column shows the port from which passengers embarked on their Titanic voyage. The three possible values are C (Cherbourg), Q (Queenstown), or S (Southampton). The distribution suggests that most passengers embarked from Southampton, while a smaller percentage embarked from Cherbourg and Queenstown.

Data Visualizations

PassengerId

200 400 600 800

Survived

0 0.2 0.4 0.6 0.8

Distribution of Pclass

0 1.25 1.50 1.75 2.00 2.25 2.50 2.75

Distribution of Age

10 20 30 40 50 60 70

Distribution of SibSp

Distribution of Parch

Flag

Use via API · Built with Gradio · Settings

25°C Rain

Search

File Explorer

Google Chrome

WhatsApp

Telegram

Discord

LinkedIn

25

ENG IN

15:44 26-07-2025

6. ****Embarked****: The Embarked column shows the port from which passengers embarked on their Titanic voyage. The three possible values are C (Cherbourg), Q (Queenstown), or S (Southampton). The distribution suggests that most passengers embarked from Southampton, while a smaller percentage embarked from Cherbourg and Queenstown.



Home

EDA_LLM-Automation

AI-Powered EDA with Mistral

LLM-Powered Exploratory Data Analysis

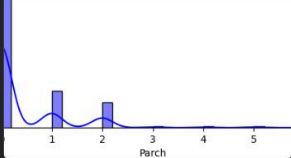

127.0.0.1:7860

Sign in

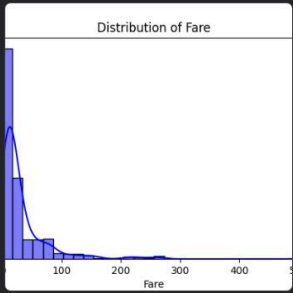
frequently (7 times). However, keep in mind that these insights may be affected by duplicate entries in some columns like Ticket and Name.

6. ****Embarked****: The Embarked column shows the port from which passengers embarked on their Titanic voyage. The three possible values are C (Cherbourg), Q (Queenstown), or S (Southampton). The distribution suggests that most passengers embarked from Southampton, while a smaller percentage embarked from Cherbourg and Queenstown.

Data Visualizations



Distribution of Fare



Correlation Heatmap

	Survived	Relatives	Age	SibSp	Parch	Fare
Survived	1.00	-0.04	0.03	-0.06	-0.00	0.01
Relatives	-0.34	1.00	-0.06	-0.04	0.08	0.26
Age	0.34	-0.34	1.00	0.08	0.02	-0.55
SibSp	-0.06	-0.34	0.21	1.00	-0.17	0.10
Parch	-0.04	0.08	-0.37	-0.17	1.00	0.16
Fare	0.01	0.26	-0.55	0.10	0.16	1.00

Flag

Use via API · Built with Gradio · Settings

25°C Rain

Search

ENG IN

15:44 26-07-2025