

llm-model-with-eda-mistral-1

June 25, 2025

```
[10]: import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
```

```
[2]: url = r"C:\Users\Bhavesh Govind Bhork\Downloads\titanic_ dataset_final.csv"
```

```
[3]: df = pd.read_csv(url)
```

```
[4]: print(df)
```

	PassengerId	Survived	Pclass	\
0	1	0	3	
1	2	1	1	
2	3	1	3	
3	4	1	1	
4	5	0	3	
--	
886	887	0	2	
887	888	1	1	
888	889	0	3	
889	890	1	1	
890	891	0	3	

	Name	Sex	Age	SibSp	\
0	Braund, Mr. Owen Harris	male	22.0	1	
1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	
2	Heikkinen, Miss. Laina	female	26.0	0	
3	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	
4	Allen, Mr. William Henry	male	35.0	0	
--	
886	Montvila, Rev. Juozas	male	27.0	0	
887	Graham, Miss. Margaret Edith	female	19.0	0	
888	Johnston, Miss. Catherine Helen "Carrie"	female	NaN	1	
889	Behr, Mr. Karl Howell	male	26.0	0	
890	Dooley, Mr. Patrick	male	32.0	0	

	Parch	Ticket	Fare	Cabin	Embarked
0	0	A/5 21171	7.2500	NaN	S

1	0	PC 17599	71.2833	C85	C
2	0	STON/O2. 3101282	7.9250	NaN	S
3	0	113803	53.1000	C123	S
4	0	373450	8.0500	NaN	S
..
886	0	211536	13.0000	NaN	S
887	0	112053	30.0000	B42	S
888	2	W./C. 6607	23.4500	NaN	S
889	0	111369	30.0000	C148	C
890	0	370376	7.7500	NaN	Q

[891 rows x 12 columns]

[5]: `print(df.describe())`

	PassengerId	Survived	Pclass	Age	SibSp \
count	891.000000	891.000000	891.000000	714.000000	891.000000
mean	446.000000	0.383838	2.308642	29.699118	0.523008
std	257.353842	0.486592	0.836071	14.526497	1.102743
min	1.000000	0.000000	1.000000	0.420000	0.000000
25%	223.500000	0.000000	2.000000	20.125000	0.000000
50%	446.000000	0.000000	3.000000	28.000000	0.000000
75%	668.500000	1.000000	3.000000	38.000000	1.000000
max	891.000000	1.000000	3.000000	80.000000	8.000000

	Parch	Fare
count	891.000000	891.000000
mean	0.381594	32.204208
std	0.806057	49.693429
min	0.000000	0.000000
25%	0.000000	7.910400
50%	0.000000	14.454200
75%	0.000000	31.000000
max	6.000000	512.329200

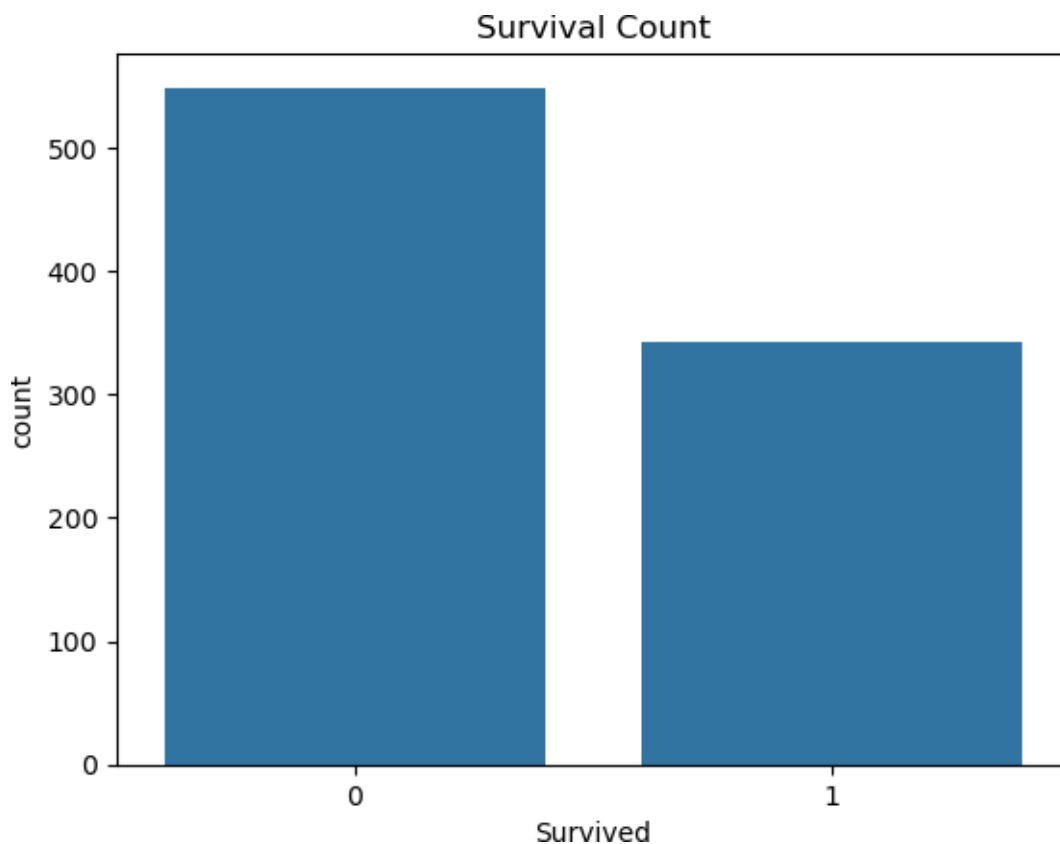
[6]: `#missing value`
`print("\nMissing values:\n", df.isnull().sum())`

Missing values:

PassengerId	0
Survived	0
Pclass	0
Name	0
Sex	0
Age	177
SibSp	0
Parch	0

```
Ticket      0
Fare        0
Cabin      687
Embarked    2
dtype: int64
```

```
[7]: #survival rate visu
sns.countplot(x='Survived', data=df)
plt.title("Survival Count")
plt.show()
```



```
[11]: %pip install ollama
import ollama
```

```
Requirement already satisfied: ollama in c:\users\bhavesht govind
bhork\anaconda3\lib\site-packages (0.5.1)
Requirement already satisfied: httpx>=0.27 in c:\users\bhavesht govind
bhork\anaconda3\lib\site-packages (from ollama) (0.27.0)
Requirement already satisfied: pydantic>=2.9 in c:\users\bhavesht govind
bhork\anaconda3\lib\site-packages (from ollama) (2.11.7)
Requirement already satisfied: anyio in c:\users\bhavesht govind
```

bhork\anaconda3\lib\site-packages (from httpx>=0.27->ollama) (4.2.0)
 Requirement already satisfied: certifi in c:\users\bhavesesh govind
 bhork\anaconda3\lib\site-packages (from httpx>=0.27->ollama) (2024.8.30)
 Requirement already satisfied: httpcore==1.* in c:\users\bhavesesh govind
 bhork\anaconda3\lib\site-packages (from httpx>=0.27->ollama) (1.0.2)
 Requirement already satisfied: idna in c:\users\bhavesesh govind
 bhork\anaconda3\lib\site-packages (from httpx>=0.27->ollama) (3.7)
 Requirement already satisfied: sniffio in c:\users\bhavesesh govind
 bhork\anaconda3\lib\site-packages (from httpx>=0.27->ollama) (1.3.0)
 Requirement already satisfied: h11<0.15,>=0.13 in c:\users\bhavesesh govind
 bhork\anaconda3\lib\site-packages (from httpcore==1.*->httpx>=0.27->ollama)
 (0.14.0)
 Requirement already satisfied: annotated-types>=0.6.0 in c:\users\bhavesesh govind
 bhork\anaconda3\lib\site-packages (from pydantic>=2.9->ollama) (0.6.0)
 Requirement already satisfied: pydantic-core==2.33.2 in c:\users\bhavesesh govind
 bhork\anaconda3\lib\site-packages (from pydantic>=2.9->ollama) (2.33.2)
 Requirement already satisfied: typing-extensions>=4.12.2 in c:\users\bhavesesh
 govind bhork\anaconda3\lib\site-packages (from pydantic>=2.9->ollama) (4.14.0)
 Requirement already satisfied: typing-inspection>=0.4.0 in c:\users\bhavesesh
 govind bhork\anaconda3\lib\site-packages (from pydantic>=2.9->ollama) (0.4.1)
 Note: you may need to restart the kernel to use updated packages.

```

[13]: import ollama
import requests

def is_ollama_running(url="http://localhost:11434"):
    try:
        response = requests.get(url)
        return response.status_code == 200
    except Exception:
        return False

def generate_insights(df_summary):
    prompt = f"Analyze the dataset summary and provide insights:
    s\n\n{df_summary}"
    try:
        response = ollama.chat(model="mistral", messages=[{"role": "user",
        s"content": prompt}])
        return response['message']['content']
    except Exception as e:
        return f"Error: Could not connect to Ollama. Please ensure Ollama is
        srunning. Details: {e}"

# Generate AI Insights
summary = df.describe().to_string()
if is_ollama_running():
    insights = generate_insights(summary)
  
```

```

    print("\n AI-Generated Insights:\n", insights)
else:
    print("Ollama server is not running. Please download and start Ollama from_
https://ollama.com/download")

```

Ollama server is not running. Please download and start Ollama from
<https://ollama.com/download>

```

[14]: %pip install gradio
import gradio as gr

def eda_analysis(file):
    df = pd.read_csv(file)
    summary = df.describe().to_string()
    insights = generate_insights(summary)
    return insights

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

demo.launch()

```

Requirement already satisfied: gradio in c:\users\bhavesh govind
bhork\anaconda3\lib\site-packages (5.34.2)
Requirement already satisfied: aiofiles<25.0,>=22.0 in c:\users\bhavesh govind
bhork\anaconda3\lib\site-packages (from gradio) (24.1.0)
Requirement already satisfied: anyio<5.0,>=3.0 in c:\users\bhavesh govind
bhork\anaconda3\lib\site-packages (from gradio) (4.2.0)
Requirement already satisfied: fastapi<1.0,>=0.115.2 in c:\users\bhavesh govind
bhork\anaconda3\lib\site-packages (from gradio) (0.115.13)
Requirement already satisfied: ffmpeg in c:\users\bhavesh govind
bhork\anaconda3\lib\site-packages (from gradio) (0.6.0)
Requirement already satisfied: gradio-client==1.10.3 in c:\users\bhavesh govind
bhork\anaconda3\lib\site-packages (from gradio) (1.10.3)
Requirement already satisfied: groovy~=0.1 in c:\users\bhavesh govind
bhork\anaconda3\lib\site-packages (from gradio) (0.1.2)
Requirement already satisfied: httpx>=0.24.1 in c:\users\bhavesh govind
bhork\anaconda3\lib\site-packages (from gradio) (0.27.0)
Requirement already satisfied: huggingface-hub>=0.28.1 in c:\users\bhavesh
govind bhork\anaconda3\lib\site-packages (from gradio) (0.33.0)
Requirement already satisfied: jinja2<4.0 in c:\users\bhavesh govind
bhork\anaconda3\lib\site-packages (from gradio) (3.1.4)
Requirement already satisfied: markupsafe<4.0,>=2.0 in c:\users\bhavesh govind
bhork\anaconda3\lib\site-packages (from gradio) (2.1.3)
Requirement already satisfied: numpy<3.0,>=1.0 in c:\users\bhavesh govind
bhork\anaconda3\lib\site-packages (from gradio) (1.26.4)
Requirement already satisfied: orjson~=3.0 in c:\users\bhavesh govind

bhork\anaconda3\lib\site-packages (from gradio) (3.10.18)
 Requirement already satisfied: packaging in c:\users\bhavesesh govind
 bhork\appdata\roaming\python\python312\site-packages (from gradio) (25.0)
 Requirement already satisfied: pandas<3.0,>=1.0 in c:\users\bhavesesh govind
 bhork\anaconda3\lib\site-packages (from gradio) (2.2.2)
 Requirement already satisfied: pillow<12.0,>=8.0 in c:\users\bhavesesh govind
 bhork\anaconda3\lib\site-packages (from gradio) (10.4.0)
 Requirement already satisfied: pydantic<2.12,>=2.0 in c:\users\bhavesesh govind
 bhork\anaconda3\lib\site-packages (from gradio) (2.11.7)
 Requirement already satisfied: pydub in c:\users\bhavesesh govind
 bhork\anaconda3\lib\site-packages (from gradio) (0.25.1)
 Requirement already satisfied: python-multipart>=0.0.18 in c:\users\bhavesesh
 govind bhork\anaconda3\lib\site-packages (from gradio) (0.0.20)
 Requirement already satisfied: pyyaml<7.0,>=5.0 in c:\users\bhavesesh govind
 bhork\anaconda3\lib\site-packages (from gradio) (6.0.1)
 Requirement already satisfied: ruff>=0.9.3 in c:\users\bhavesesh govind
 bhork\anaconda3\lib\site-packages (from gradio) (0.12.0)
 Requirement already satisfied: safehttpx<0.2.0,>=0.1.6 in c:\users\bhavesesh
 govind bhork\anaconda3\lib\site-packages (from gradio) (0.1.6)
 Requirement already satisfied: semantic-version~=2.0 in c:\users\bhavesesh govind
 bhork\anaconda3\lib\site-packages (from gradio) (2.10.0)
 Requirement already satisfied: starlette<1.0,>=0.40.0 in c:\users\bhavesesh govind
 bhork\anaconda3\lib\site-packages (from gradio) (0.46.2)
 Requirement already satisfied: tomlkit<0.14.0,>=0.12.0 in c:\users\bhavesesh
 govind bhork\anaconda3\lib\site-packages (from gradio) (0.13.3)
 Requirement already satisfied: typer<1.0,>=0.12 in c:\users\bhavesesh govind
 bhork\anaconda3\lib\site-packages (from gradio) (0.16.0)
 Requirement already satisfied: typing-extensions~=4.0 in c:\users\bhavesesh govind
 bhork\anaconda3\lib\site-packages (from gradio) (4.14.0)
 Requirement already satisfied: uvicorn>=0.14.0 in c:\users\bhavesesh govind
 bhork\anaconda3\lib\site-packages (from gradio) (0.34.3)
 Requirement already satisfied: fsspec in c:\users\bhavesesh govind
 bhork\anaconda3\lib\site-packages (from gradio-client==1.10.3->gradio)
 (2024.6.1)
 Requirement already satisfied: websockets<16.0,>=10.0 in c:\users\bhavesesh govind
 bhork\anaconda3\lib\site-packages (from gradio-client==1.10.3->gradio) (15.0.1)
 Requirement already satisfied: idna>=2.8 in c:\users\bhavesesh govind
 bhork\anaconda3\lib\site-packages (from anyio<5.0,>=3.0->gradio) (3.7)
 Requirement already satisfied: sniffio>=1.1 in c:\users\bhavesesh govind
 bhork\anaconda3\lib\site-packages (from anyio<5.0,>=3.0->gradio) (1.3.0)
 Requirement already satisfied: certifi in c:\users\bhavesesh govind
 bhork\anaconda3\lib\site-packages (from httpx>=0.24.1->gradio) (2024.8.30)
 Requirement already satisfied: httpcore==1.* in c:\users\bhavesesh govind
 bhork\anaconda3\lib\site-packages (from httpx>=0.24.1->gradio) (1.0.2)
 Requirement already satisfied: h11<0.15,>=0.13 in c:\users\bhavesesh govind
 bhork\anaconda3\lib\site-packages (from httpcore==1.*->httpx>=0.24.1->gradio)
 (0.14.0)
 Requirement already satisfied: filelock in c:\users\bhavesesh govind

bhork\anaconda3\lib\site-packages (from huggingface-hub>=0.28.1->gradio)
 (3.13.1)
 Requirement already satisfied: requests in c:\users\bhavesesh govind
 bhork\anaconda3\lib\site-packages (from huggingface-hub>=0.28.1->gradio)
 (2.32.3)
 Requirement already satisfied: tqdm>=4.42.1 in c:\users\bhavesesh govind
 bhork\anaconda3\lib\site-packages (from huggingface-hub>=0.28.1->gradio)
 (4.66.5)
 Requirement already satisfied: python-dateutil>=2.8.2 in c:\users\bhavesesh govind
 bhork\appdata\roaming\python\python312\site-packages (from
 pandas<3.0,>=1.0->gradio) (2.9.0.post0)
 Requirement already satisfied: pytz>=2020.1 in c:\users\bhavesesh govind
 bhork\anaconda3\lib\site-packages (from pandas<3.0,>=1.0->gradio) (2024.1)
 Requirement already satisfied: tzdata>=2022.7 in c:\users\bhavesesh govind
 bhork\anaconda3\lib\site-packages (from pandas<3.0,>=1.0->gradio) (2023.3)
 Requirement already satisfied: annotated-types>=0.6.0 in c:\users\bhavesesh govind
 bhork\anaconda3\lib\site-packages (from pydantic<2.12,>=2.0->gradio) (0.6.0)
 Requirement already satisfied: pydantic-core==2.33.2 in c:\users\bhavesesh govind
 bhork\anaconda3\lib\site-packages (from pydantic<2.12,>=2.0->gradio) (2.33.2)
 Requirement already satisfied: typing-inspection>=0.4.0 in c:\users\bhavesesh
 govind bhork\anaconda3\lib\site-packages (from pydantic<2.12,>=2.0->gradio)
 (0.4.1)
 Requirement already satisfied: click>=8.0.0 in c:\users\bhavesesh govind
 bhork\anaconda3\lib\site-packages (from typer<1.0,>=0.12->gradio) (8.1.7)
 Requirement already satisfied: shellingham>=1.3.0 in c:\users\bhavesesh govind
 bhork\anaconda3\lib\site-packages (from typer<1.0,>=0.12->gradio) (1.5.4)
 Requirement already satisfied: rich>=10.11.0 in c:\users\bhavesesh govind
 bhork\anaconda3\lib\site-packages (from typer<1.0,>=0.12->gradio) (13.7.1)
 Requirement already satisfied: colorama in c:\users\bhavesesh govind
 bhork\appdata\roaming\python\python312\site-packages (from
 click>=8.0.0->typer<1.0,>=0.12->gradio) (0.4.6)
 Requirement already satisfied: six>=1.5 in c:\users\bhavesesh govind
 bhork\appdata\roaming\python\python312\site-packages (from python-
 dateutil>=2.8.2->pandas<3.0,>=1.0->gradio) (1.17.0)
 Requirement already satisfied: markdown-it-py>=2.2.0 in c:\users\bhavesesh govind
 bhork\anaconda3\lib\site-packages (from rich>=10.11.0->typer<1.0,>=0.12->gradio)
 (2.2.0)
 Requirement already satisfied: pygments<3.0.0,>=2.13.0 in c:\users\bhavesesh
 govind bhork\appdata\roaming\python\python312\site-packages (from
 rich>=10.11.0->typer<1.0,>=0.12->gradio) (2.19.1)
 Requirement already satisfied: charset-normalizer<4,>=2 in c:\users\bhavesesh
 govind bhork\anaconda3\lib\site-packages (from requests->huggingface-
 hub>=0.28.1->gradio) (3.3.2)
 Requirement already satisfied: urllib3<3,>=1.21.1 in c:\users\bhavesesh govind
 bhork\anaconda3\lib\site-packages (from requests->huggingface-
 hub>=0.28.1->gradio) (2.2.3)
 Requirement already satisfied: mdurl~=0.1 in c:\users\bhavesesh govind
 bhork\anaconda3\lib\site-packages (from markdown-it-

py>=2.2.0->rich>=10.11.0->typer<1.0,>=0.12->gradio) (0.1.0)
Note: you may need to restart the kernel to use updated packages.
* Running on local URL: http://127.0.0.1:7860
* To create a public link, set `share=True` in `launch()`.
<IPython.core.display.HTML object>

[14]:

```
[15]: import gradio as gr
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
import ollama

# Function to Perform EDA and Generate Visualizations
def eda_analysis(file_path):
    df = pd.read_csv(file_path)

    # Fill missing values with median for numeric columns
    for col in df.select_dtypes(include=['number']).columns:
        df[col].fillna(df[col].median(), inplace=True)

    # Fill missing values with mode for categorical columns
    for col in df.select_dtypes(include=['object']).columns:
        df[col].fillna(df[col].mode()[0], inplace=True)

    # Data Summary
    summary = df.describe(include='all').to_string()

    # Missing Values
    missing_values = df.isnull().sum().to_string()

    # Generate AI Insights
    insights = generate_ai_insights(summary)

    # Generate Data Visualizations
    plot_paths = generate_visualizations(df)

    return f"\n Data Loaded Successfully!\n\n Summary:\n{summary}\n\n Missing_
sValues:\n{missing_values}\n\n AI Insights:\n{insights}", plot_paths

# AI-Powered Insights using Mistral-7B (Ollama)
def generate_ai_insights(df_summary):
    prompt = f"Analyze the dataset summary and provide insights:
s\n\n{df_summary}"
    response = ollama.chat(model="mistral", messages=[{"role": "user",
s"content": prompt}])
```



```

    return response['message']['content']

# Function to Generate Data Visualizations
def generate_visualizations(df):
    plot_paths = []

    # Histograms for Numeric Columns
    for col in df.select_dtypes(include=['number']).columns:
        plt.figure(figsize=(6,4))
        sns.histplot(df[col], bins=30, kde=True, color="blue")
        plt.title(f"Distribution of {col}")
        path = f"{col}_distribution.png"
        plt.savefig(path)
        plot_paths.append(path)
        plt.close()

    # Correlation Heatmap (only numeric columns)
    numeric_df = df.select_dtypes(include=['number'])
    if not numeric_df.empty:
        plt.figure(figsize=(8,5))
        sns.heatmap(numeric_df.corr(), annot=True, cmap='coolwarm', fmt=".2f",
        linewidths=0.5)
        plt.title("Correlation Heatmap")
        path = "correlation_heatmap.png"
        plt.savefig(path)
        plot_paths.append(path)
        plt.close()

    return plot_paths

# Gradio Interface
demo = gr.Interface(
    fn=eda_analysis,
    inputs=gr.File(type="filepath"),
    outputs=[gr.Textbox(label="EDA Report"), gr.Gallery(label="Data_
    Visualizations")],
    title=" LLM-Powered Exploratory Data Analysis (EDA)",
    description="Upload any dataset CSV file and get automated EDA insights_
    with AI-powered analysis and visualizations."
)

# Launch the Gradio App (local only, avoids antivirus issues)
demo.launch()


```

* Running on local URL: <http://127.0.0.1:7861>

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

<IPython.core.display.HTML object>

[15]:



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

3. The majority of passengers (577) are male, while only 314 are female. This could indicate a gender imbalance on the ship.

4. The average age of passengers is approximately 29 years old, with a standard deviation of around 13 years. The minimum age is 0.42, suggesting that an infant was on board.

5. On average, each passenger had about 0.5 siblings aboard and about 0.4 family members traveling with them (parents or children).

6. The average ticket fare is approximately 32.20, with a significant range between the minimum of 0 and maximum of 512.33. The Cabin feature has missing values for most passengers, which could indicate that not all passengers had assigned cabins.

7. Most passengers (691) embarked from Southampton, 646 passengers embarked from Cherbourg, and only 46 passengers embarked from Queenstown.

Missing Values:
PassengerId: 0

Data Visualizations

passengerid

Survived

Distribution of Pclass

Distribution of Age

Distribution of SibSp

Distribution of Parch

Flag

Use via API · Built with Gradio · Settings

10