Jupyter Notebook Example to PDF

October 26, 2023

[ ]:

[ ]:
```
# Exploratory Data Analysis (EDA)

# Welcome to this Jupyter Notebook, where we will explore the fascinating world␣
↪of EDA in data science. EDA is the first step in any data analysis process,␣
↪allowing us to understand our data before diving into more complex tasks.

# Let's get started!
```

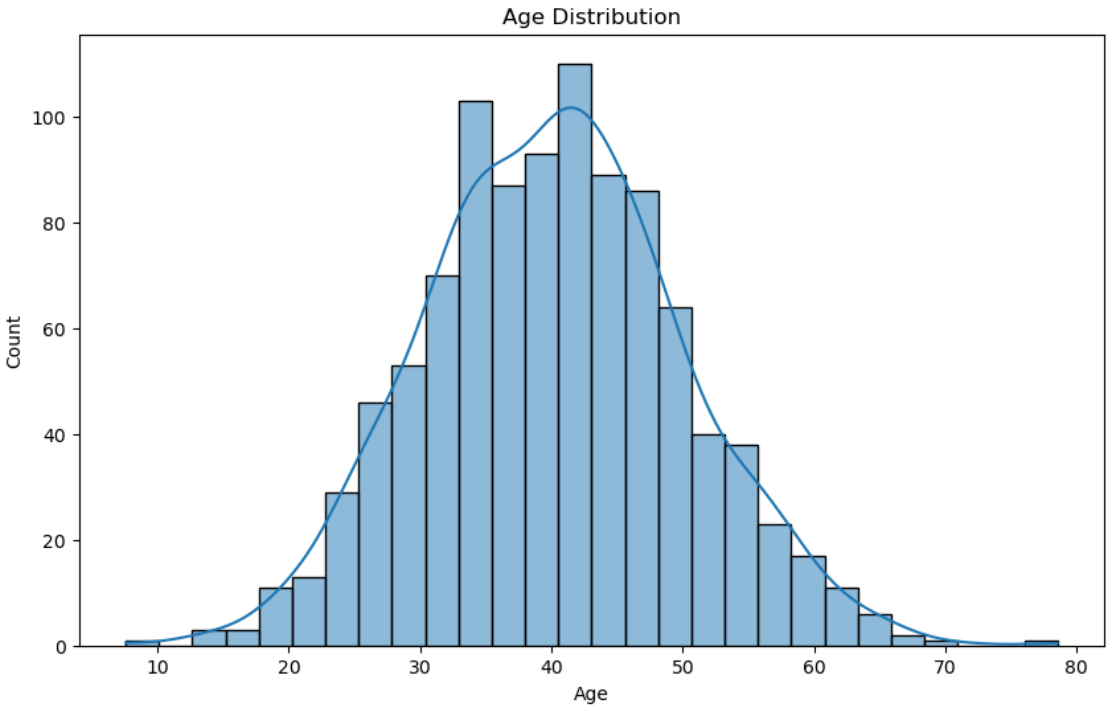
[ ]:
```
# Import Libraries
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
```

[ ]:
```
# Generating synthetic data
np.random.seed(42)
n_samples = 1000

age = np.random.normal(40, 10, n_samples)
income = np.random.normal(50000, 10000, n_samples)

data = pd.DataFrame({'age': age, 'income': income})
```

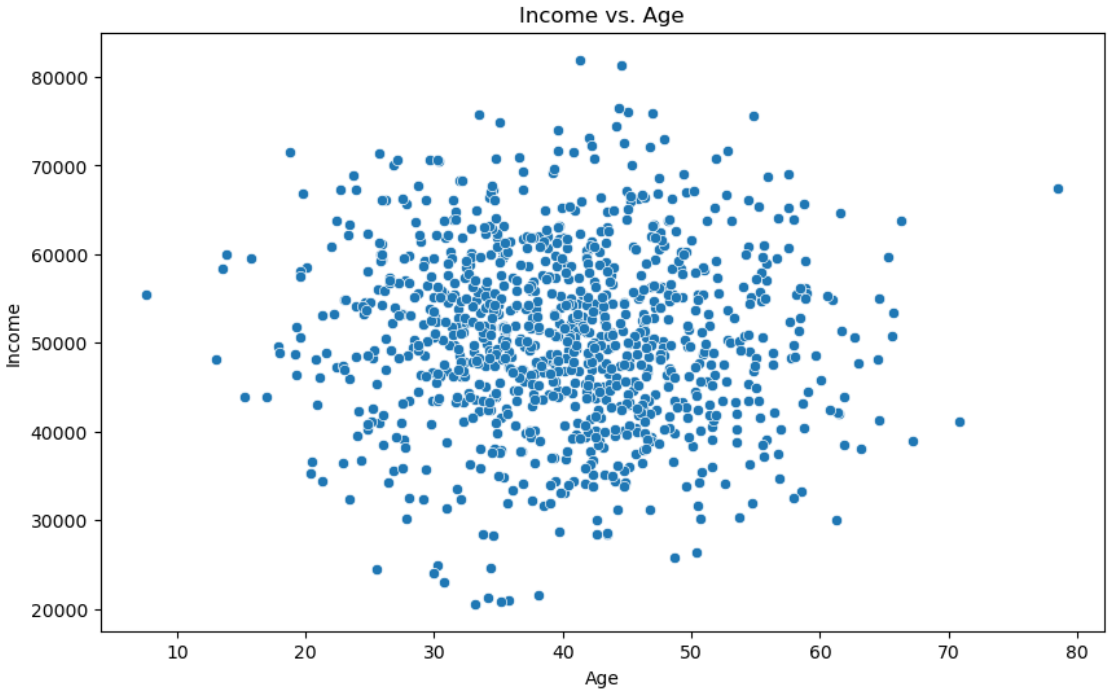
[ ]:
```
# Display the first 5 rows of the dataset
data.head()

# Check for missing values
missing_values = data.isnull().sum()

# Summary statistics
summary_stats = data.describe()
```

[ ]:
```
plt.figure(figsize=(10, 6))
sns.histplot(data['age'], kde=True)
plt.title('Age Distribution')
plt.xlabel('Age')
```

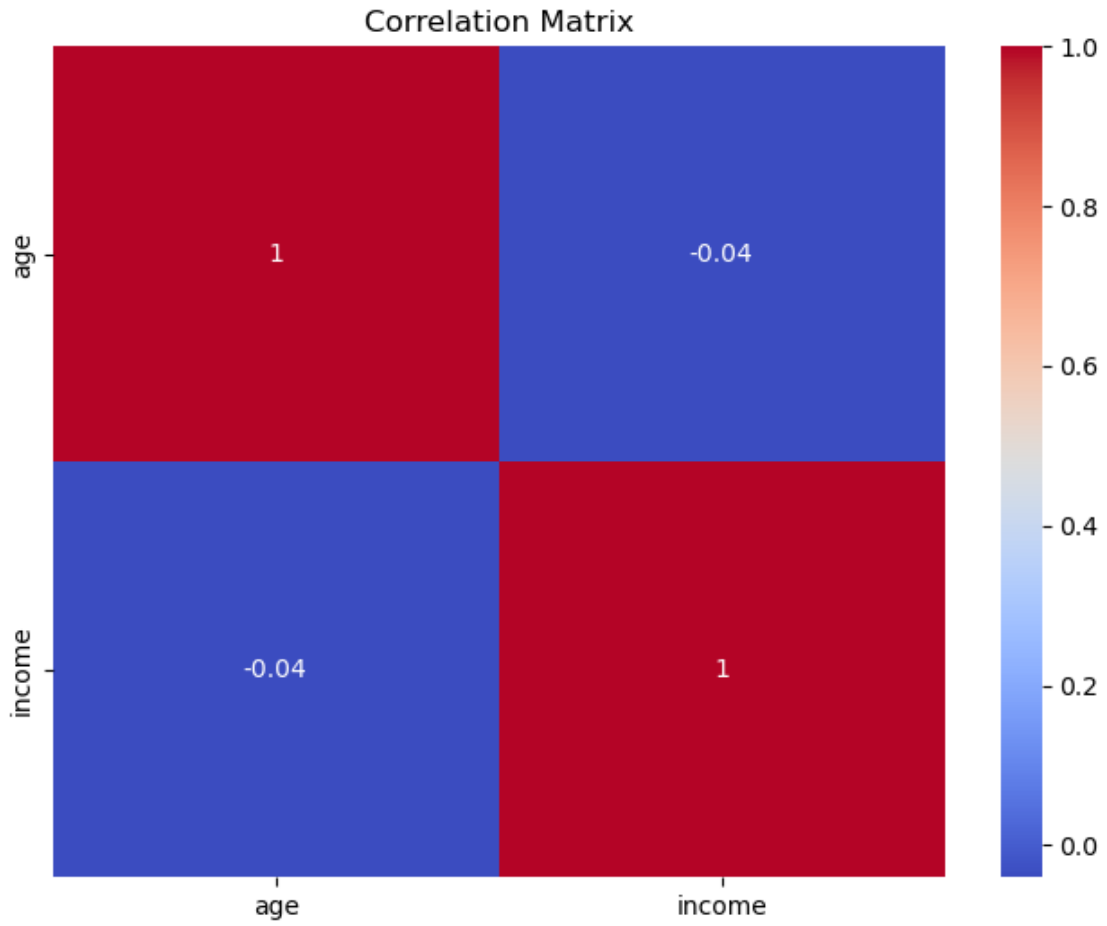
```
plt.ylabel('Count')
plt.show()
```



[ ]:
```
plt.figure(figsize=(10, 6))
sns.scatterplot(x='age', y='income', data=data)
plt.title('Income vs. Age')
plt.xlabel('Age')
plt.ylabel('Income')
plt.show()
```



[ ]:
```
correlation_matrix = data.corr()

plt.figure(figsize=(8, 6))
sns.heatmap(correlation_matrix, annot=True, cmap='coolwarm')
plt.title('Correlation Matrix')
plt.show()
```



[ ]:
```
# Conclusion

# In this Jupyter Notebook, we explored the basics of Exploratory Data Analysis.
 ↪ We generated synthetic data, performed data exploration, and created␣
 ↪visualizations to better understand our data.

# EDA is a crucial step in the data science process, as it helps us identify␣
 ↪patterns, outliers, and relationships within the data. These insights are␣
 ↪essential for making informed decisions and building accurate models.

# Remember, this notebook is just the tip of the iceberg when it comes to EDA.␣
 ↪You can dive deeper into more advanced techniques and real datasets in your␣
 ↪data science journey.

# Feel free to extend and customize this notebook with more data analysis and␣
 ↪visualization techniques to suit your needs.
```

```
# Happy data exploring!
```

[ ]:
```
# This notebook generates synthetic data for demonstration purposes, which you␣
↪can further customize to fit your data science project's needs.
```

[ ]:
```

```