

pandasai_cvs

April 29, 2024

```
[ ]: from dotenv import load_dotenv

load_dotenv(override=True)
```

```
[ ]: True
```

```
[ ]: import pandas as pd

data = pd.read_csv("data.csv")
data.head()
```

```
[ ]:
```

	Name	Age	Gender	City	Occupation	Income
0	Wang Lei	32	Male	Beijing	Teacher	10000
1	Li Wei Jia	23	Female	Shanghai	Doctor	20000
2	Zheng Xiao Gang	45	Male	Guangzhou	Policeman	10000
3	Huang Xiao Ning	28	Male	Shenzhen	Programmer	15000
4	He Li Na	35	Female	Chengdu	Lawyer	25000

```
[ ]: import os
from langchain_groq.chat_models import ChatGroq

llm = ChatGroq(
    model_name="mixtral-8x7b-32768",
    api_key=os.environ["GROQ_API_KEY"],
)
```

```
[ ]: from pandasai import SmartDataframe

df = SmartDataframe(data, config={"llm": llm})
```

?

```
[ ]: df.chat("What are the top 5 cities for average income?")
```

```
[ ]:
```

	Income
City	
Qingdao	40000.0
Shenyang	37500.0
Xi'an	32500.0

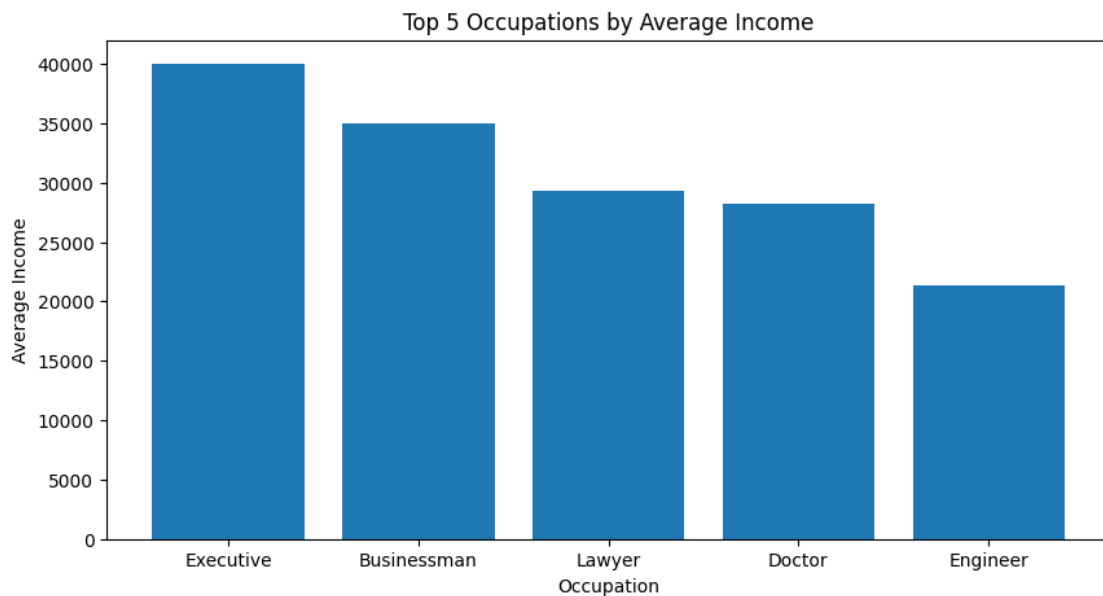
```
Kunming    31500.0
Chengdu    26600.0
```

5

```
[ ]: df.chat(
    "Calculate the average income for each occupation and then rank the top 5_
    ↳occupations from highest to lowest"
)
```

<string>:2: FutureWarning: The default value of numeric_only in DataFrameGroupBy.mean is deprecated. In a future version, numeric_only will default to False. Either specify numeric_only or select only columns which should be valid for the function.

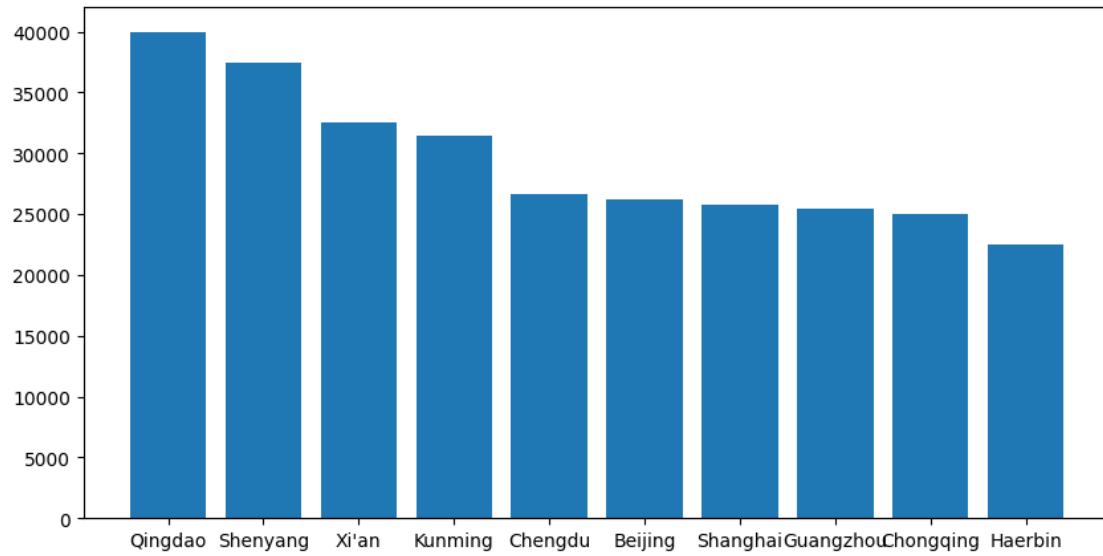
```
[ ]: '/Users/chi/Desktop/quartz/content/study/exports/charts/temp_chart.png'
```



10

```
[ ]: df.chat(
    "First, sort the cities by average income from high to low, then create a_
    ↳bar chart displaying the top 10 cities by average income"
)
```

```
[ ]: '/Users/chi/Desktop/quartz/content/study/exports/charts/temp_chart.png'
```



5

```
[ ]: df.chat(
    "First, calculate the number of people in each occupation, and then create_
    ↪ a pie chart for the top 5 occupations by count"
)
```

```
[ ]: '/Users/chi/Desktop/quartz/content/study/exports/charts/temp_chart.png'
```

