

Correlation matrix

Creating a correlation matrix on Google Cloud Vertex AI:

1. Data Preparation:

- Data was loaded on BigQuery, and was cleaned by handling all the missing values, string values and any other inconsistencies.
- Choose columns that you want to include in the correlation matrix and drop the ones you do not need. Ensure they are numerical or can be converted into numerical format.

The screenshot displays the Google Cloud BigQuery interface. The top section shows a table preview with columns: Row, totalCharges, gender, SeniorCitizen, Dependents, Partner, tenure, PhoneService, MultipleLines, and InternetService. The table contains 13 rows of data. Below the table, there is a 'Job history' section with a 'REFRESH' button. At the bottom, there is a 'Launcher' section with various notebook and console options.

Row	totalCharges	gender	SeniorCitizen	Dependents	Partner	tenure	PhoneService	MultipleLines	InternetService
1	73.45	1	0	0	0	4	1	0	0
2	78.9	1	0	0	0	4	1	0	0
3	348.8	1	0	0	0	18	1	0	0
4	1149.65	1	0	0	0	54	1	0	0
5	141.6	0	0	0	0	8	1	0	0
6	1396.0	1	1	0	0	68	1	0	0
7	211.95	0	0	0	0	11	1	0	0
8	198.6	1	0	0	0	10	1	0	0
9	417.7	1	0	0	0	21	1	0	0
10	198.7	1	0	0	0	10	1	0	0
11	185.4	1	0	0	0	10	1	0	0
12	63.6	0	0	0	0	3	1	0	0
13	275.7	1	0	0	0	16	1	0	0

Results per page: 50 1 - 50 of 7043

Job history REFRESH

Launcher

Notebook

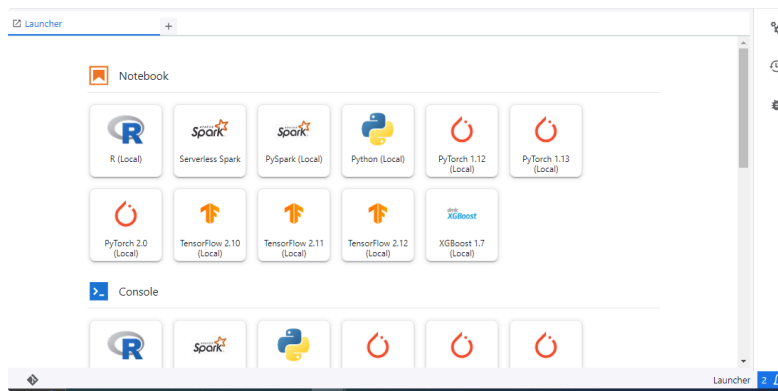
- R (Local)
- Serverless Spark
- PySpark (Local)
- Python (Local)
- PyTorch 1.12 (Local)
- PyTorch 1.13 (Local)
- PyTorch 2.0 (Local)
- TensorFlow 2.10 (Local)
- TensorFlow 2.11 (Local)
- TensorFlow 2.12 (Local)
- XGBoost 1.7 (Local)

Console

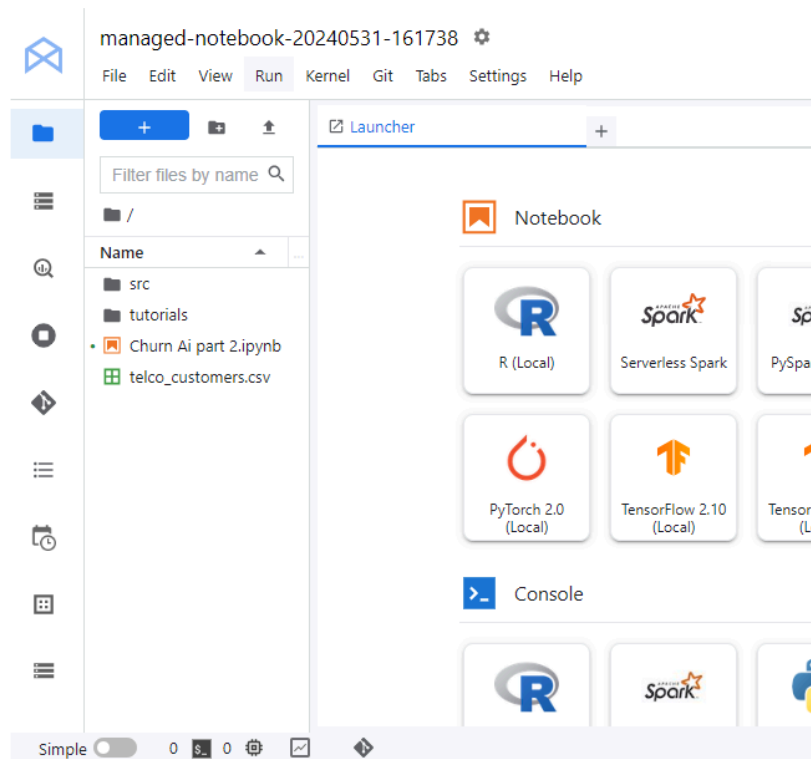
- R
- Spark
- Python
- PyTorch
- PyTorch
- PyTorch

2. Model Training and Output Visualization:

- Enable the necessary APIs like Vertex AI, Cloud Storage, and BigQuery if you're using them.
- Go to the Vertex AI section in Google Cloud Console and open a new notebook (Python Local).



- Upload the cleaned dataset to Google Cloud Storage, by dragging and dropping it.



- Create or import a Pandas dataframe, compute the correlation matrix using the `.corr()` method, and visualize it with a heatmap to represent the correlations effectively.
- Write the Python code to calculate the correlation matrix. You can use libraries like NumPy or Pandas for computation. If you get errors try downloading the libraries before getting them.
- For visualization use tools like Matplotlib, and Seaborn to visualize the correlation matrix. You can generate plots directly within your Python code or save them as image files.

```
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt

df = pd.read_csv('telco_customers.csv')
correlation_matrix = df.corr()

print(correlation_matrix) # Add closing parenthesis

plt.figure(figsize=(10, 8))
sns.heatmap(correlation_matrix, annot=True, cmap='coolwarm', vmin=-1, vmax=1)
plt.title('Correlation Matrix')
plt.show()
```

- Optionally, you can save the correlation matrix and its visualization to Google Cloud Storage or any other desired storage location for future reference.

