Task 3 (10 points)

Describe the correlation. The answer should be comprehensive. (5 points)

The practical example of application in cybersecurity with data and the python code are mandatory. (5 points)

Place your answer into the task3.pdf file. The source code and the data should be included in the resulting pdf file. No additional files are required.

Understanding connections between different pieces of information can be really important, especially when it comes to spotting potential issues or threats. It's like putting puzzle pieces together, but with data instead of physical pieces.

In this line of work, we often deal with large amounts of data, like logs or traffic records, and the goal is to find patterns or irregularities that could indicate something fishy going on. That's where analyzing the relationships between different variables or features comes into play.

It's kind of like looking at a group of friends and trying to figure out who influences each other's behavior the most. Some friends might have a stronger impact on certain behaviors, while others might have little to no influence. By understanding these connections, we can better predict how the group might act in different situations.

The process starts by looking at data samples, each with its own set of characteristics or features. Some samples are labeled as "good" and others as "not so good," based on specific criteria. Then, a technique called "correlation analysis" is used to identify which features are most strongly connected to the "not so good" label.

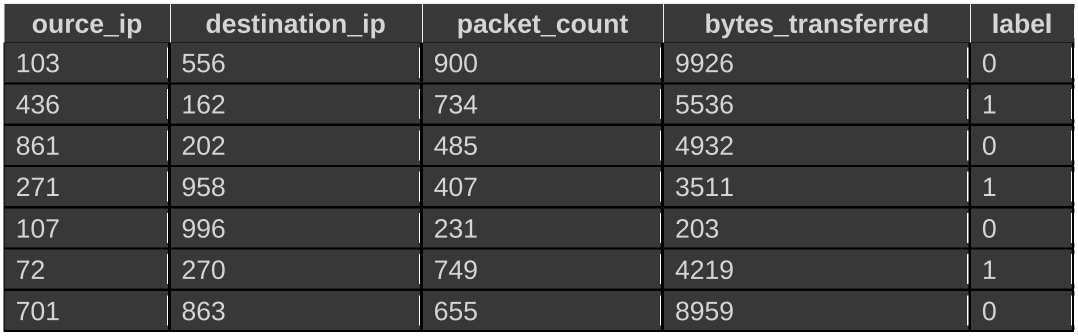
It's like finding the ringleaders in a group of troublemakers. The features that have a strong positive relationship with the "not so good" label are the ones that need close monitoring, because they're more likely to be associated with problematic behavior.

Once these connections are understood, models can be built to predict which new samples might be "not so good" based on their feature values. It's like having a crystal ball that can tell us which new data points might be a cause for concern, based on their individual characteristics.

Of course, it's not an exact science, and care must be taken not to jump to conclusions too quickly. Sometimes, the connections identified might be coincidental or misleading, so the models need to be constantly validated and fine-tuned based on new information.

But overall, this whole process of analyzing correlations and building predictive models based on those connections has been really useful in helping to stay ahead of potential threats and keep things running smoothly.

Data:



ource_ip	destination_ip	packet_count	bytes_transferred	label
103	556	900	9926	0
436	162	734	5536	1
861	202	485	4932	0
271	958	407	3511	1
107	996	231	203	0
72	270	749	4219	1
701	863	655	8959	0

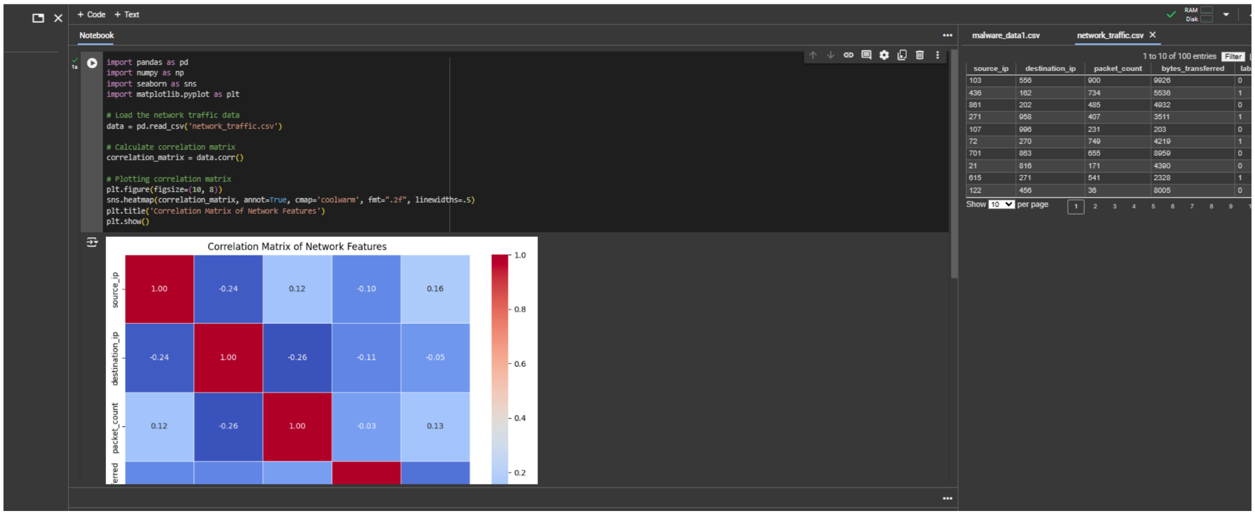
ource_ip	destination_ip	packet_count	bytes_transferred	label
21	816	171	4390	0

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

# Load the network traffic data
data = pd.read_csv('network_traffic.csv')

# Calculate correlation matrix
correlation_matrix = data.corr()

# Plotting correlation matrix
plt.figure(figsize=(10, 8))
sns.heatmap(correlation_matrix, annot=True, cmap='coolwarm', fmt=".2f", linewidths=.5)
plt.title('Correlation Matrix of Network Features')
plt.show()
```



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

# Load the network traffic data
data = pd.read_csv('network_traffic.csv')

# Calculate correlation matrix
correlation_matrix = data.corr()

# Plotting correlation matrix
plt.figure(figsize=(10, 8))
sns.heatmap(correlation_matrix, annot=True, cmap='coolwarm', fmt=".2f", linewidths=.5)
plt.title('Correlation Matrix of Network Features')
plt.show()
```

1 to 10 of 100 entries

source_ip	destination_ip	packet_count	bytes_transferred	lab
103	556	900	9926	0
436	162	734	5536	1
861	202	485	4932	0
271	958	407	3511	1
107	996	231	203	0
72	270	749	4219	1
701	863	655	8959	0
21	816	171	4390	0
615	271	541	2328	1
122	456	36	8005	0

Show 10 per page 1 2 3 4 5 6 7 8 9