Python Scripts

Python Correlation Heatmap:

Displayed the *relationship between job title, department, salary, and bonus percentage,* allowing for insights into whether certain job titles or departments were associated with higher compensation packages.

```
# The following code to create a dataframe and remove duplicated rows is always executed and acts as a preamble for your script:


# dataset = pandas.DataFrame(Bonus %, Annual Salary, Job Title, Department)

# dataset = dataset.drop_duplicates()


# Paste or type your script code here:

import pandas as pd

import matplotlib.pyplot as plt

import seaborn as sns


# Drop duplicate rows

dataset = dataset.drop_duplicates()


# Convert categorical columns into numerical categories

dataset['Job Title'] = pd.factorize(dataset['Job Title'])[0]

dataset['Department'] = pd.factorize(dataset['Department'])[0]


# Select relevant columns

df = dataset[['Job Title', 'Department', 'Annual Salary', 'Bonus %']]


# Calculate the correlation matrix
```

```
corr_matrix = df.corr()


# Plot the heatmap of the correlation matrix
plt.figure(figsize=(12, 10))


# Create the heatmap
ax = sns.heatmap(corr_matrix, annot=True, cmap='tab20b', fmt='.2f',
                 annot_kws={"size": 32},  # Font size for annotations
                 cbar_kws={"shrink": .8})  # Adjust color bar size


# Adjust the size of the color bar legend text
colorbar = ax.collections[0].colorbar
colorbar.ax.tick_params(labelsize=20)  # Font size for color bar ticks
colorbar.ax.set_ylabel(colorbar.ax.get_ylabel(), fontsize=20)  # Font size for color bar label


# Title and axis labels with increased font size
plt.title("Correlation Between Job Title, Department, Salary, and Bonus %", fontsize=22)
plt.xlabel('Variables', fontsize=32)
plt.ylabel('Variables', fontsize=32)


# Adjust tick label font sizes
plt.xticks(fontsize=18)
plt.yticks(fontsize=18)


# Display the plot
plt.show()
```

Histogram:

Visualized the *distribution of annual salaries,* showing the frequency of different salary ranges within the company.

```
# The following code to create a dataframe and remove duplicated rows is always executed and acts as a preamble for your script:


# dataset = pandas.DataFrame(Annual Salary)

# dataset = dataset.drop_duplicates()


# Paste or type your script code here:

import matplotlib.pyplot as plt

import numpy as np

x = np.random.normal(dataset["Annual Salary"])


plt.hist(x, color = "purple", edgecolor = "white")

plt.title("Distribution of Annual Salary", fontsize = 15)

plt.xlabel("Annual Salary", fontsize = 15)

plt.ylabel("Frequency", fontsize = 15)


plt.show()
```