Jupyter Notebook

Jupyter is a great tool for auditors because it allows you to create, store, and save workpapers similar to OneNote for notetaking.

Let's download a few base packages to get started. These packages help us import and analyze data as well as create visualizations

In [ ]:
```
!pip3 install pandas pyod matplotlib numpy seaborn bokeh plotly fbprophet
```

In [ ]:
```
# Let's import data - 5 Million Credit Card Transactions
```

In [ ]:
```
import pandas as pd

df_cc = pd.read_csv("5m CC Records.csv")

df_cc.head()

# what is a 'df' dataframe?
```

In [ ]:
```
# Let's import data - 5 Million Sales Records
```

In [ ]:
```
import pandas as pd

df_sales = pd.read_csv("5m Sales Records.csv")

df_sales.head()
```

In [ ]:
```
# Let's analyze data
```

In [ ]:
```
df_sales.describe()
```

In [ ]:
```
# Let's improve our formatting
```

In [ ]:
```
pd.options.display.float_format = '{:.2f}'.format
df_sales.describe()
```

In [ ]:
```
df_sales.describe()[['Units Sold', 'Total Cost']]
```

In [ ]:
```
df_cc.head()
```

In [ ]:
```
#Analyzing transactions by type
```

In [ ]:
```
df_cc['Description'].value_counts()
```

In [ ]:
```
# Find our most expensive bills
df_cc[df_cc['Description']=='ATM'].sort_values(by="Withdrawls", ascending=False
```

In [ ]:
```
import pandas as pd
```

```
df_new = pd.read_csv("5m CC Records.csv")
df_new.to_csv("New File Name")
```

In [ ]:
```
# Let's plot some data
```

In [ ]:
```
import pandas as pd
import matplotlib.pyplot as plt


# reading the database
data = pd.read_csv("5m Sales Records.csv")

# hostogram of total_bills
plt.hist(data['Total Revenue'])

plt.title("Histogram")

plt.ticklabel_format(style='plain')
# Adding the legends
plt.show()
```

In [ ]:
```
#View withdrawals over time
import pandas as pd_a

data = pd_a.read_csv("5m CC Records.csv")
data = data.apply(lambda x: x.str.replace(',', ''))
pd_a.options.display.float_format = '{:.2f}'.format

data["Withdrawls"] = pd_a.to_numeric(data["Withdrawls"])
data = data.sort_values(by="Date")

data.set_index('Date')['Withdrawls'].plot();
```

In [ ]:
```
# importing packages
import seaborn as sns
import matplotlib.pyplot as plt
import pandas as pd


# reading the database
data = pd.read_csv("5m Sales Records.csv")

# draw lineplot
sns.lineplot(x="Sales Channel", y="Total Profit", data=data)

# setting the title using Matplotlib
plt.title('Sales Channel and Total Profit')

plt.show()
```

In [ ]:
```
import plotly.express as px
fig = px.histogram(df_sales, x='Total Cost')

fig.show()
```

In [ ]:
```
#create a box plot
```

```
fig = px.box(df_sales, y='Total Cost')
fig.show()
```

In [ ]: