CERTIFICATION - DATA ANALYST - PRODUCT SALES

Data Analyst Professional Practical Exam Submission

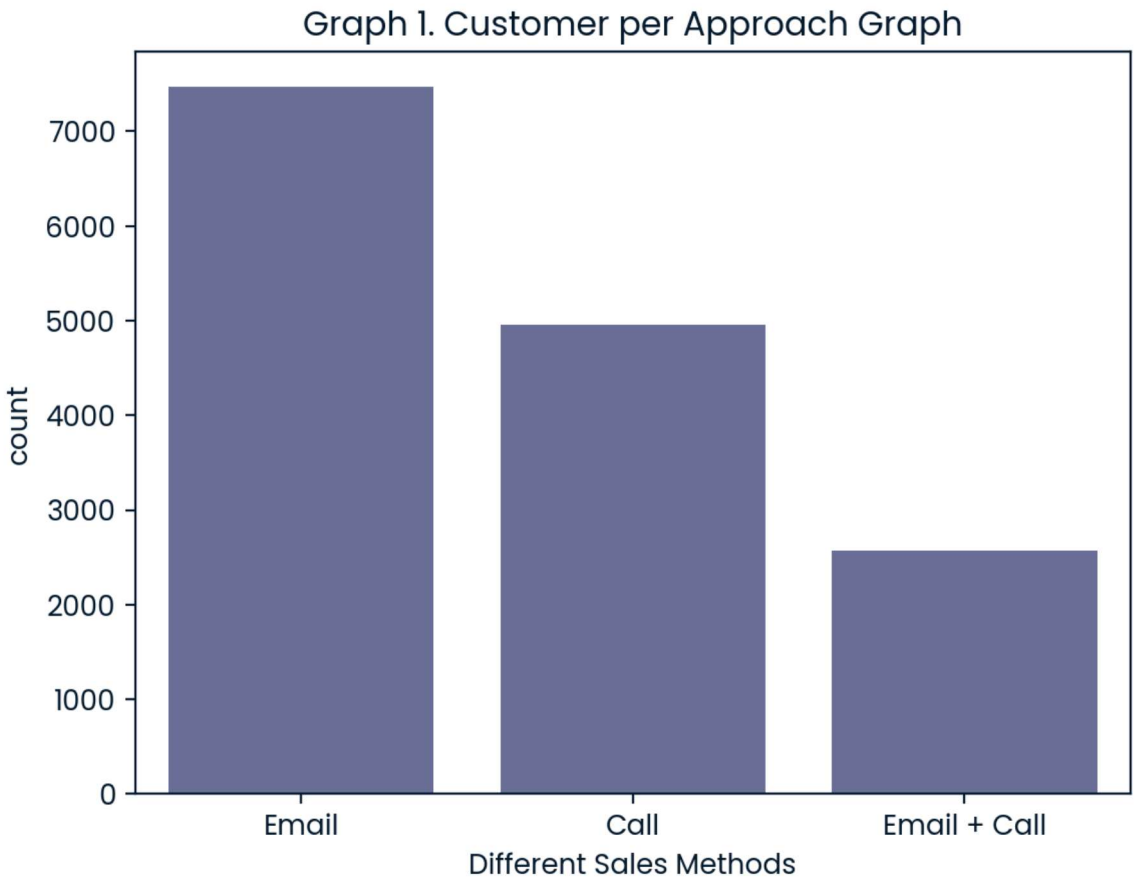
You can use any tool that you want to do your analysis and create visualizations. Use this template to write up your summary for submission.

You can use any markdown formatting you wish. If you are not familiar with Markdown, read the **Markdown Guide** before you start.

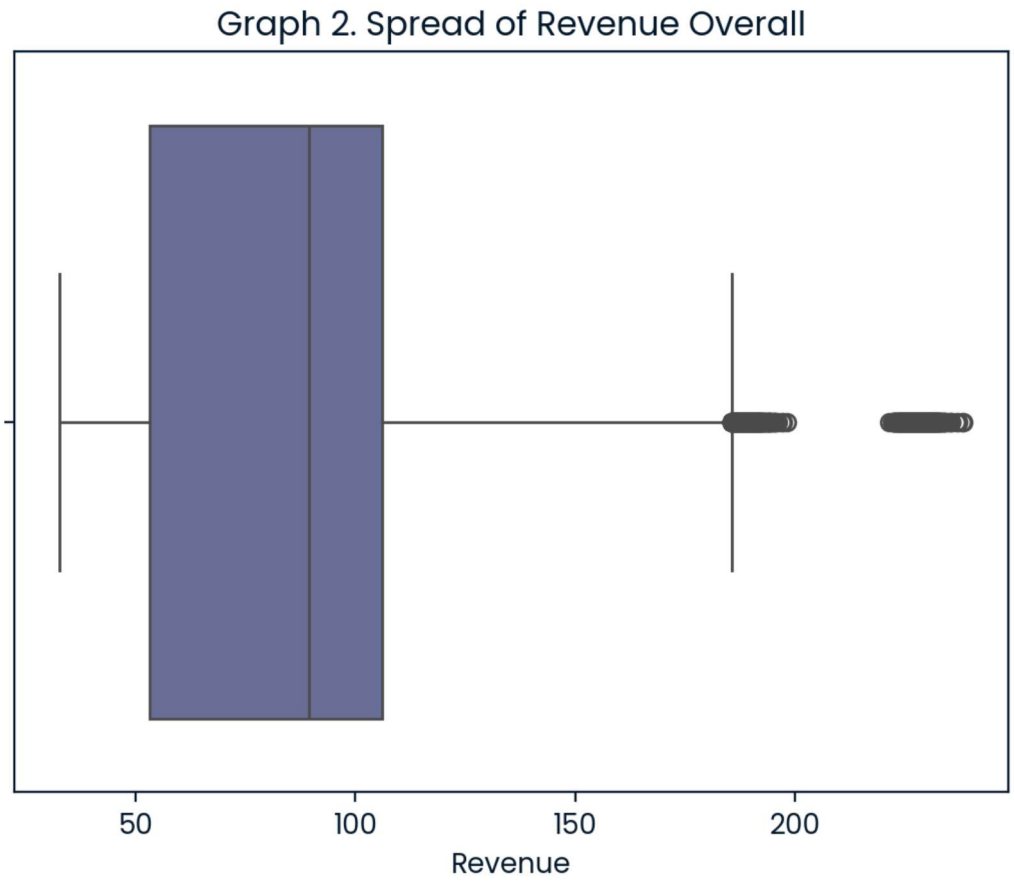
```
# Q1 : How many customers were there for each approach?
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns

# importing the file to analyze
ps = pd.read_csv("product_sales.csv")

# cleaning our data before analysis
ps['sales_method'] = ps['sales_method'].replace({
    "em + call":"Email + Call",
    "email":"Email"
})
# Since the revenue distribution is right-skewed, I imputed missing values with the
median rather than the mean, as the median is a more reliable central measure in
skewed data
ps['revenue'] = ps['revenue'].astype(float) #making sure all values are indeed
numeric before cleaning
ps['revenue'] = ps['revenue'].fillna(ps['revenue'].median())
ps = ps[ps['years_as_customer'] <=39]

# answering first question of : how many customers were there for each approach
num_of_cust = ps['sales_method'].value_counts()
print(num_of_cust)
# im using order here to have have better storytelling, this graph clearly showcases
that Email is by far the most popular and Email + Call is the least
order = ps['sales_method'].value_counts().index
sns.countplot(data=ps,x='sales_method', order=order)
plt.title("Graph 1. Customer per Approach Graph")
plt.xlabel("Different Sales Methods")
```

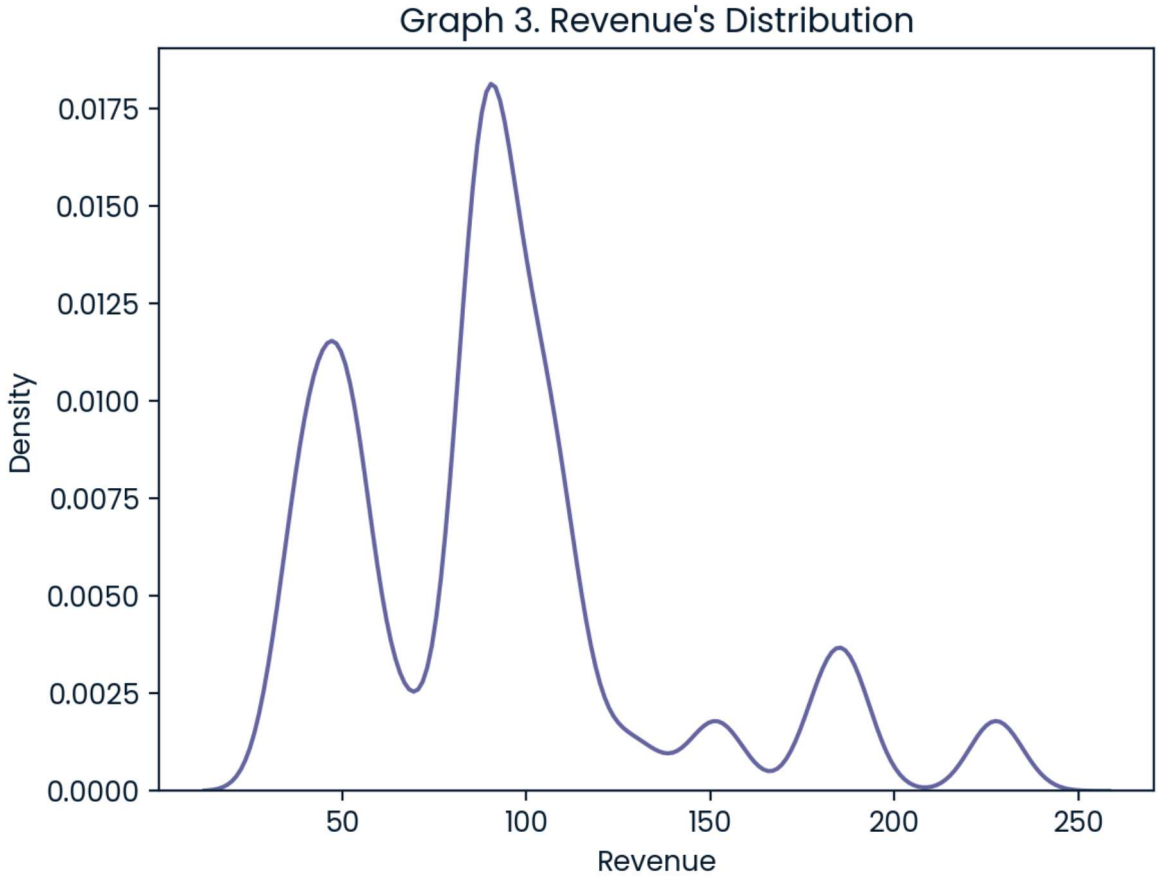
```
Email           7465
Call            4961
Email + Call    2572
Name: sales_method, dtype: int64
```

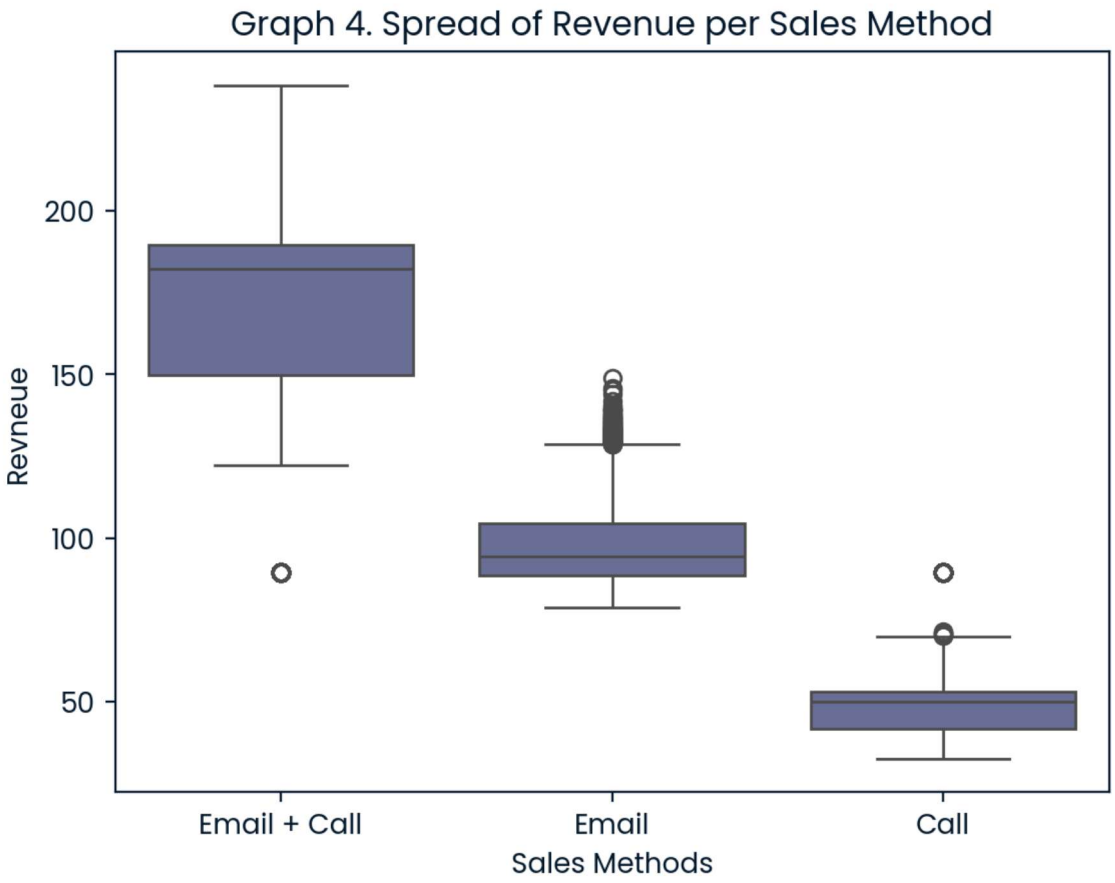
```
Text(0.5, 0, 'Different Sales Methods')
```


Graph 1. Customer per Approach Graph
7000
6000
5000
4000
3000
2000
1000
0
count
Email
Call
Email + Call
Different Sales Methods

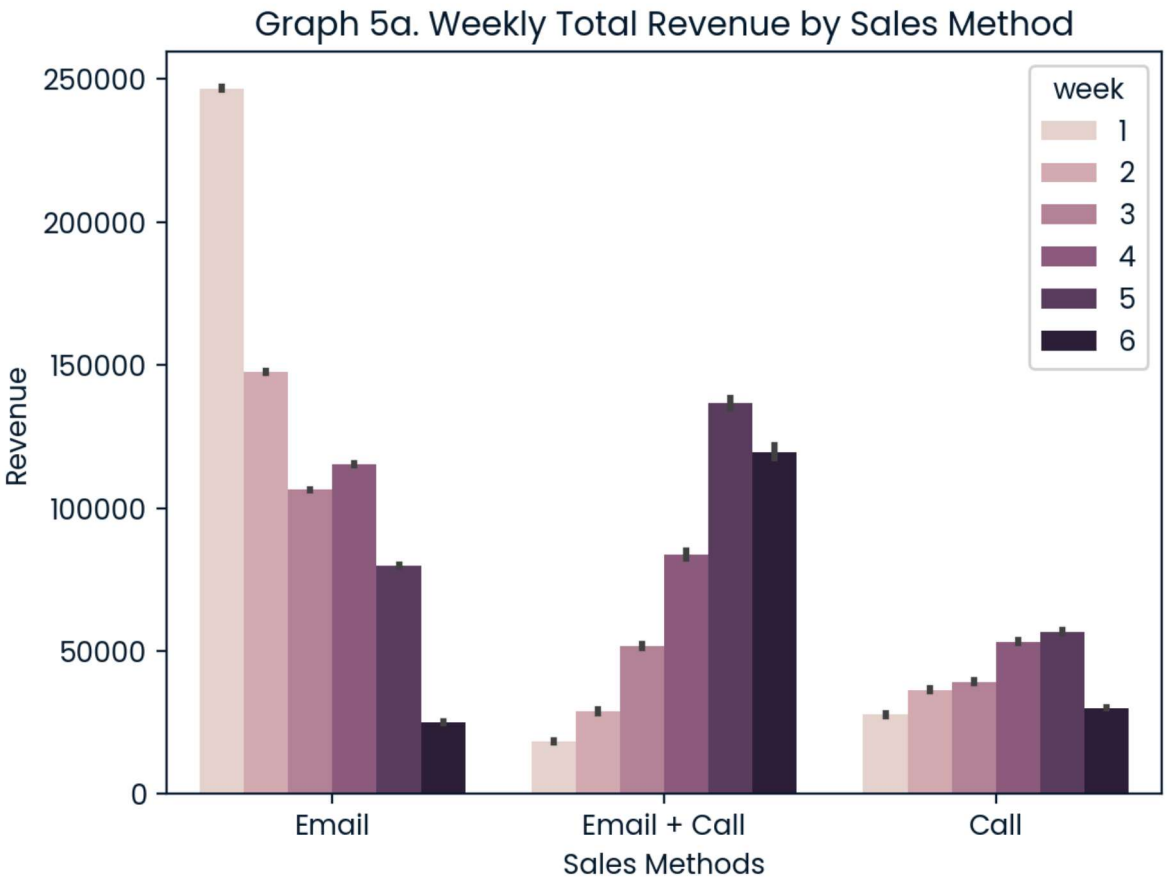
```
# Q2 : What does the spread of the revenue look like overall? And for each method?
# boxplot to visualize spread of revenue overall and revenue per sales method

# revenue overall
sns.boxplot(data=ps, x='revenue')
plt.title("Graph 2. Spread of Revenue Overall")
plt.xlabel("Revenue")
plt.show()
# I wan't to visualize distrbiution here as well for further analysis
sns.kdeplot(data=ps, x='revenue')
plt.title("Graph 3. Revenue's Distribution")
plt.xlabel("Revenue")
plt.show()
# we can see that the distribution is right skew'd which is tied back to the
cleaning section of Q1

# revenue per method
per_method_order = ( ps.groupby('sales_method')['revenue']
                     .sum()
                     .index
                     .sort_values(ascending=False)
)
sns.boxplot(data=ps, x='sales_method', y='revenue', order=per_method_order)
plt.title("Graph 4. Spread of Revenue per Sales Method")
plt.xlabel("Sales Methods")
plt.ylabel("Revneue")
plt.show()
```

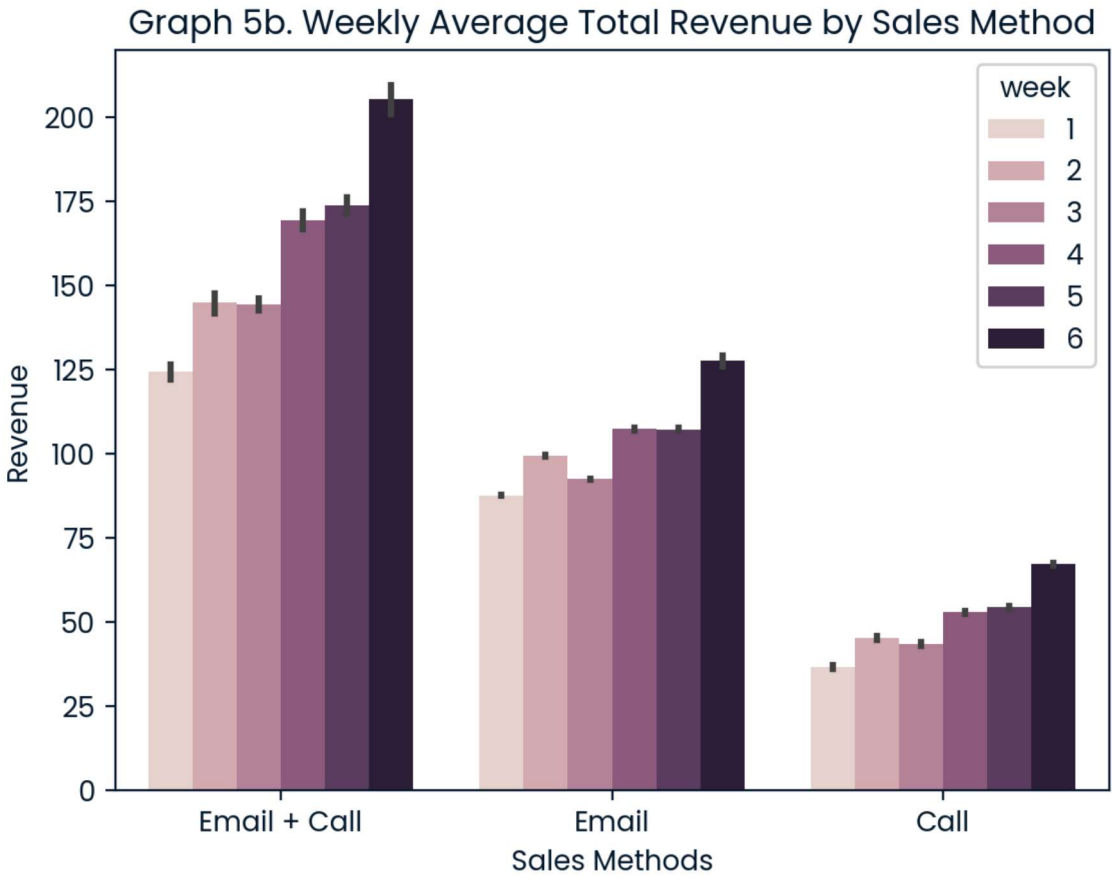

Graph 2. Spread of Revenue Overall
50
100
150
200
Revenue


Graph 3. Revenue's Distribution
0.0175
0.0150
0.0125
0.0100
0.0075
0.0050
0.0025
0.0000
Density
50
100
150
200
250
Revenue


Graph 4. Spread of Revenue per Sales Method
Revneue
200
150
100
50
Email + Call
Email
Call
Sales Methods

```
# Q3 : Was there any difference in revenue over time for each of the methods?

# visualization (TOTAL)
diff_over_weeks =( ps.groupby('sales_method')['revenue']
                      .sum()
                      .sort_values(ascending=False)
                      .index
)
sns.barplot(data=ps, x='sales_method', y='revenue', hue='week',
order=diff_over_weeks, estimator=sum)
plt.title("Graph 5a. Weekly Total Revenue by Sales Method")
plt.xlabel("Sales Methods")
plt.ylabel("Revenue")
plt.show()

# visualization (AVERAGE)
diff_over_weeks =( ps.groupby('sales_method')['revenue']
                      .mean()
                      .sort_values(ascending=False)
                      .index
)
sns.barplot(data=ps, x='sales_method', y='revenue', hue='week',
order=diff_over_weeks)
plt.title("Graph 5b. Weekly Average Total Revenue by Sales Method")
plt.xlabel("Sales Methods")
plt.ylabel("Revenue")
plt.show()
```


Graph 5a. Weekly Total Revenue by Sales Method
week
1
2
3
4
5
6
Revenue
250000
200000
150000
100000
50000
0
Email
Email + Call
Call
Sales Methods


Graph 5b. Weekly Average Total Revenue by Sales Method
week
1
2
3
4
5
6
Revenue
200
175
150
125
100
75
50
25
0
Email + Call
Email
Call
Sales Methods

📝 Task List

Your written report should include written text summaries and graphics of the following:

- Data validation:
  - Describe validation and cleaning steps for every column in the data
- Exploratory Analysis:
  - Include two different graphics showing single variables only to demonstrate the characteristics of data
  - Include at least one graphic showing two or more variables to represent the relationship between features
  - Describe your findings
- Definition of a metric for the business to monitor
  - How should the business use the metric to monitor the business problem
  - Can you estimate initial value(s) for the metric based on the current data
- Final summary including recommendations that the business should undertake

Start writing report here..

✅ When you have finished...

- Publish your Workspace using the option on the left
- Check the published version of your report:
  - Can you see everything you want us to grade?
  - Are all the graphics visible?
- Review the grading rubric. Have you included everything that will be graded?
- Head back to the **Certification Dashboard** to submit your practical exam report and record your presentation