Instructions

- This homework assignment is worth 65 points.
- Please submit a **.ipynb** file to Blackboard.
- **Please strive for clarity and organization.**
- **Due Date: September 22, 2023 by 11:59 pm.**

Exercise 1

(8 points) An online movie streaming company has a business problem of growing customer churn. That is, subscription customers are canceling their subscriptions to join the competitor. Create an analytics plan that could be used to address this business problem.

Exercise 2

(8 points) Although their sales are reasonable, an online fashion retailer is struggling to generate the volume of sales that they had originally hoped for when launching their site. Create an analytics plan that could be used to address this business problem.

Exercise 3

The sales of some products in your company in the past few years have been increasing and the company wants to build a marketing strategy for them. Using Python answer the following:

(a) (4 points) Using the `pandas` read the csv file called `sales.csv` and create a data-frame called `sales`.

(b) (3 points) Report the number of observations and variables in the `sales` data-frame.

(c) (4 points) Using the appropriate Python commands, create a frequency table of the `Product` variable. Report the least popular `Product`.

(d) (4 points) Using the appropriate Python commands, compute the summary statistics of `Revenue`.

(e) (4 points) Using the appropriate Python commands, create a frequency table of the `Product type` variable. Report the most popular `Product type`.

(f) (5 points) Compute the total `Revenue` by `Year`. What year has the highest revenue?

(g) (5 points) Compute the total `Revenue` by `Order method type`. What channel has the highest revenue?

Exercise 4

(5 points) What is exploratory analysis and why it is important.

Exercise 5

(5 points) Suppose you work for a cloud storage company and you analyze the amount of content uploaded every month. During early November, you notice a spike in picture uploads. What could be the cause of the spike? Answering the this question is what type of data analytics?

Exercise 6

Let the data `x` be given by

```
import numpy as np

x = np.array([1, 8, 2, 6, 3, 8, 5, 5, 5, 5])
```

Note that X_1 denotes the first element of `x`, which is 1. **In Python**, answer the following:

(a) (3 points) Create the array `x`.

(b) (3 points) Find the mean of `x`.

(c) (3 points) Compute the standard deviation of `x`.

(d) (3 points) Find $\sum_{i=1}^{10} \log(X_i)$ using the `sum` and `log` functions.

(e) (3 points) Report the IQR of `x`.