Faculty of Science

# Assignment 2

Weight: 15% final grade
To be completed by an individual or in pairs (groups of 2)
Due date: April 1, 2022 11:59pm

**Overview**
Data processing is a useful and common task in computer programming. The goal of data processing is to prepare raw data for a specific task. In this assignment, you will be working with an open data source to process, summarize, and visualize some data. It is recommended that you use IntelliJ but you can use any IDE you are comfortable with. Git and Gradle are also recommended but not required.

**Instructions**
For this assignment you are expected to work individually or with a partner. Please download the CSV file (*airline_safety.csv*) from Canvas under Assignment 2 and complete the tasks outlined below.

1. **Process the Data (5%) –** Write a Java program to read in the .csv data file. In your Java program, add a column to the data that represents the *total number of incidents between 1985 and 2014*. Convert the data into an XML document, and write the XML data to a file called ***converted_airline_safety.xml*.**

2. **Summarize the Data (5%)** – Using the CSV or XML data, calculate the following information:
   a. The minimum, maximum, and average values for **each column**.
   b. The **average** number of incidents between 1985 – 1999 **across all airlines**.
   c. The **average** number of incidents between 2000 – 2014 **across all airlines**.

   Write these statistics to an XML file called ***airline_summary_statistic.xml***. The format of your text file should resemble the example in **Figure 1**. Replace column_name, min_val, max_val, and avg_val with your calculated values. Make sure you include stats for each column in the data set.

3. **Create a Chart using JavaFX (5%)** – Using JavaFX, display a vertical or horizontal **bar chart** that displays *2 bars for each airline*. The first bar should indicate the number of fatal incidents between 1985 – 1999, and the second bar should indicate the number of fatal incidents between 2000 – 2014. Please see the following link for an example of a bar chart showing more than 1 bar per category (https://docs.oracle.com/javafx/2/charts/bar-chart.htm).

summary.xml

```xml
<Summary>
    <Stat>
        <Name>column_name</Name>
        <Min>min_val</Min>
        <Max>max_val</Max>
        <Avg></Avg>
    </Stat>
    ...
    <Stat>
        <Name>avg_85_99</Name>
        <Min></Min>
        <Max></Max>
        <Avg>avg_val</Avg>
    </Stat>
    <Stat>
        <Name>avg_00_14</Name>
        <Min></Min>
        <Max></Max>
        <Avg>avg_val</Avg>
    </Stat>
</Summary>
```

**Figure 1: airline_summary_statistics.xml**

**Data Source**

The CSV file for this assignment is available on Canvas under Assignment 2. You can find more information regarding this data source at the following link:

https://github.com/fivethirtyeight/data/tree/master/airline-safety

**Submission Instructions (Due on April 1, 2022 by 11:59pm)**

1. Submit a **.zip file on Canvas** containing the following files:
    a. Your project directory.
    b. The original CSV and generated XML files in the **'resources'** folder of your project.

**Grading Guidelines**

This is not a rubric, but general points to guide you. You will be graded considering:

1. Processing
    a. Correct conversion from CSV to XML
    b. Includes required information
2. Summary
    a. Correctly calculates summary statistics
    b. Writes statistics to XML file
3. JavaFX Chart
    a. Is the information on the chart accurate?
    b. Does the chart include a title, axis labels, and a legend?