

Carleton University
Department of Economics
ECON 3880, Summer 2024

Assignment 3

Due date: 11 June 2024, 11:59 pm

Professor: Firouz Fallahi

Instruction: If you haven't installed the **tidyverse** library yet, do so now.

Download and read the data set SYRflights2021.rds from Brightspace and save it in the same folder as your R Notebook file. Then read it into RStudio and call it **syr** for simplicity. This data set contains flights data from the Syracuse Airport in 2021, which is similar to the New York flights dataset that we have seen in class and has the same variables.

Please use an R Notebook to answer the following questions and submit the HTML file.

- 1 . Load the tidyverse and write a code to print the first 3 rows of dataset. What are the names of columns in this dataset?
- 2 . Use one of the functions we have learned in class to get the structure of data and understand the variables in the syr dataset. What types of data do we have in this dataset?
- 3 . Find all flights with more than 5 hours delayed arrival that flew to New York city (there are 3 airports with the codes JFK, LGA, and EWR in this city). How many flights match these two conditions?
- 4 . Use 2 different approaches that we have learned in class, to create new variables called **distance_km1** and **distance_km2** which converts the distance from miles to KM (note that 1 mile = 1.61 KM). Make sure that both variables are added to the **syr** dataset. Next, create a new dataset called **df_new** by selecting the following columns: MONTH, carrier, distance_km1, distance_km2, and all the columns that their name ends with "time". Finally show the last 4 rows of data.

Good luck!