CS 341, Spring 2025
Project 3 – Divvy Rides Data Analysis with F#
Due: Wednesday 04/09/2025 at 11:59pm

Overview

The goal of this project is to write a program in F# to perform analysis on data for Chicago Divvy bike rides.

The Divvy bike system in Chicago is a commuting service. A person may choose to become a subscriber, which costs about $145 a year, and thus gets 45 minutes free on each bike ride. Otherwise, a single ride costs $1 to unlock the bike and then $0.18 per minute for the entire ride. The anonymized data on rides is publicly available on their website: https://divvybikes.com/system-data. The data is also available via the Chicago Data Portal: https://data.cityofchicago.org/Transportation/Divvy-Trips/fg6s-gzvg/about_data.

The program begins by having the user input a filename and then proceeds to read in and analyze the data.

The Divvy Ride Data

The input file will be a text file in CSV (comma separated value) format. Each line consists of the following values, in order:

- **Trip ID:** an integer that uniquely identifies a trip, i.e. ride
- **Bike ID:** an integer that uniquely identifies the bike that was used for this trip
- **Trip Duration:** an integer that described the number of seconds that the trip lasted
- **Starting Hour:** an integer representing the hour at which the trip started (between 0 and 23, inclusive)
- **From Station ID:** an integer that represents the ID of the station from which the bike was checked out
- **To Station ID:** an integer that represents the ID of the station to which the bike was returned
- **Subscriber:** an integer that represents whether the rider is a subscriber or not: the value is 0 if they are not a subscriber and 1 if they are
- **Gender:** an integer representing the gender that the rider identifies as: the value is 0 if the gender was not specified, 1 if they identified as male, and 2 if they identified as female
- **Birth Year:** an integer representing the year in which the rider was born: the value is 0 if the birth year was not specified, and the birth year otherwise

Note that the file does *not* have column headers, only the values, which are in the order shown above. You may assume that any file entered by the user will follow this format. We reserve the right to use files outside of the ones provided to you to grade your program.

CS 341, Spring 2025
Project 3 – Divvy Rides Data Analysis with F#
Due: Wednesday 04/09/2025 at 11:59pm

The Analysis

The provided starter code includes code to open the file and input the data into a list of lists. Your job is to add the following analysis:

- Number of trips in the file
- Number of riders identifying as male in the file, along with the percentage
- Number of riders identifying as female in the file, along with the percentage
- Average age of the riders in the file
- Number of rides for each range of trip duration, along with the percentage for each
- Display a histogram of start times for the rides in the file

Number of Trips

Display the number of trips, i.e. the number of rides that are contained in the file data.

Number of Riders Identifying as Male / Female

Display the number of riders that identify as male, and the number of riders that identify as female. Also display the percentage of each, which is taken out of the total number of rides (including those for which the gender is not specified).

Average Age

Display the average age of the riders. For this calculation, do not include riders that did not specify their birth year. The function System.DateTime.Now.Year returns the current year, which can be used to calculate the age of a given rider.

Number of Rides for Each Range of Trip Duration

Display the number of rides for each of the following ranges of trip durations:

- 0-30 minutes (less than or equal to 30 minutes)
- 30-60 minutes (greater than 30 minutes, but less than or equal to 60 minutes)
- 60-120 minutes (greater than 60 minutes, but less than or equal to 120 minutes)
- >2 hours (greater than 2 hours)

The maximum trip duration is 24 hours.

Also display the percentages for each, which are taken out of the total number of rides.

Histogram of Start Times

Display a histogram of the number of rides for each starting hour, from 0 (12 AM) to 23 (11 PM). The histogram shows asterisks as the “bars” – the number of asterisks to output is the number of rides divided by 100. Also display the number of trips at the end of each line in the histogram.

CS 341, Spring 2025
Project 3 – Divvy Rides Data Analysis with F#
Due: Wednesday 04/09/2025 at 11:59pm

Sample Output

Below is sample output for the program, with user input shown in red. Note that each of the histogram lines is one line of output, although it appears to span multiple lines below.

```
Project 3: Divvy Rides Data Analysis with F#
CS 341, Spring 2025

This application allows you to analyze and visualize
information about Divvy bike rides in Chicago, such as
the number of male/female riders, the average age, etc.

Enter the name of the file with the Divvy ride data: divvy-07-01-2018.csv

Number of Trips: 11645

Number of Riders Identifying as Male: 5375 (46.15714899%)
Number of Riders Identifying as Female: 2548 (21.88063547%)

Average Age: 40.1155494

Ride Durations:
   0-30 mins: 9033 (77.56977243%)
   30-60 mins: 1600 (13.73980249%)
   60-120 mins: 757 (6.500644053%)
   > 2 hours: 255 (2.189781022%)

Histogram of Start Times:
   0: **262
   1: **204
   2: *111
   3: 66
   4: 43
   5: 45
   6: *125
   7: *179
   8: ***334
   9: *****536
   10: *******769
   11: ********873
   12: **********1128
   13: **********1067
   14: *********991
   15: **********1012
   16: *********948
   17: ******607
   18: ******605
   19: ****469
```

CS 341, Spring 2025
Project 3 – Divvy Rides Data Analysis with F#
Due: Wednesday 04/09/2025 at 11:59pm

```
   20: *****516
   21: ***369
   22: **225
   23: *161

Exiting program.

------------------------------------------------------------------------------------

Project 3: Divvy Rides Data Analysis with F#
CS 341, Spring 2025

This application allows you to analyze and visualize
information about Divvy bike rides in Chicago, such as
the number of male/female riders, the average age, etc.

Enter the name of the file with the Divvy ride data: divvy-01-2019.csv

Number of Trips: 103192

Number of Riders Identifying as Male: 80250 (77.76765641%)
Number of Riders Identifying as Female: 18864 (18.28048686%)

Average Age: 43.60038564

Ride Durations:
   0-30 mins: 98120 (95.0848903%)
   30-60 mins: 4093 (3.966392744%)
   60-120 mins: 654 (0.6337700597%)
   > 2 hours: 325 (0.3149468951%)

Histogram of Start Times:
   0: *****550
   1: ***352
   2: **242
   3: *182
   4: ***342
   5: ***********1262
   6: ****************************************4148
   7: ********************************************************************************
****8317
   8: ********************************************************************************
**********************************11211
   9: ***************************************************5139
   10: *********************************3244
   11: ****************************************4100
   12: ************************************************4852
   13: ***********************************************4727
   14: ************************************************4802
   15: ************************************************************6303
```

CS 341, Spring 2025
Project 3 – Divvy Rides Data Analysis with F#
Due: Wednesday 04/09/2025 at 11:59pm

```
   16: ********************************************************************************
*****************************10880
   17: ********************************************************************************
**************************************************13245
   18: ********************************************************************************
7885
   19: ***********************************************4813
   20: ****************************2824
   21: ******************1885
   22: ***********1199
   23: ******688

Exiting program.
```

CS 341, Spring 2025
Project 3 – Divvy Rides Data Analysis with F#
Due: Wednesday 04/09/2025 at 11:59pm

Requirements

Do not use imperative style programming: no mutable variables, no arrays, and no loops. You may write the program using recursion, tail recursion, higher-order functions, or a combination of these. Strive for efficient solutions using the techniques we have discussed in class.

It is also expected that you use good programming practices, i.e. functions, comments, consistent spacing, etc. In a 3xx class this should be obvious and not require further explanation. But to be clear, if you submit a program with no functions and no comments, you will be significantly penalized --- in fact you can expect a score of 0, even if the program produces the correct results. How many functions? How many comments? You can decide. Make reasonable decisions and you will not be penalized.

Submission

Login to Gradescope.com and look for the assignment “Project 03”. Submit your “Program.fs” to that assignment.

To encourage local testing, you will be limited to 50 submissions (submissions that result in a score of zero do not count towards this limit).

Keep in mind that any manual grading for requirements, etc. will be based on your last submission unless you select an earlier submission for grading. If you do choose to activate an earlier submission, you must do so before the deadline.

The score reported on Gradescope is only part of your final score. After the project is due, the TAs will manually review the programs for style and adherence to requirements (0-100%).

Late submissions for this project are allowed. You may turn in a project up to 3 days (72 hours) late, and will receive the following penalty on your total score:

Submission made **up to 24 hours late**: **10 point** deduction

Submission made **24-48 hours late**: **20 point** deduction

Submission made **48-72 hours late**: **30 point** deduction

No submissions will be accepted after 72 hours.

CS 341, Spring 2025
Project 3 – Divvy Rides Data Analysis with F#
Due: Wednesday 04/09/2025 at 11:59pm

Academic Integrity

All work is to be done individually — group work is not allowed.

While we encourage you to talk to your peers and learn from them, this interaction must be superficial with regards to all work submitted for grading. This means you cannot work in teams, you cannot work side-by-side, you cannot submit someone else's work (partial or complete) as your own, etc. The University's policy is available here: https://dos.uic.edu/community-standards/

In particular, note that **you are guilty of academic dishonesty if you extend or receive any kind of unauthorized assistance**. Absolutely no transfer of program code between students is permitted (paper or electronic), and you may not solicit code from family, friends, or online forums (e.g. you cannot download answers from Chegg). Other examples of academic dishonesty include emailing your program to another student, sharing your screen so that another student may copy your work, copying-pasting code from the internet, working together in a group, and allowing a tutor, TA, or another individual to write an answer for you.

Academic dishonesty is unacceptable, and penalties range from a letter grade drop to expulsion from the university; cases are handled via the official student conduct process described at the link above.