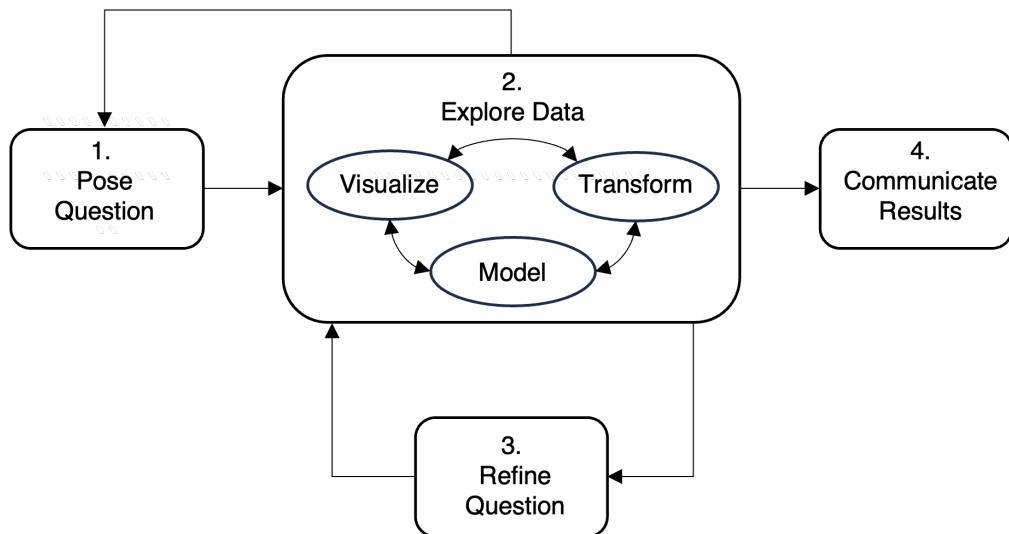


CT5102 Assignment 6 - Exploratory Data Analysis

Objective

The aim of this assignment is to use the exploratory data analysis (see below) process on a tibble that you have been allocated. Here are the guidelines:

- Pose **three questions** to explore as part of your analysis, and then generate insights from the data using the tools of the tidyverse. You may also find the package `lubridate` useful when dealing with processing dates. Any modelling activity can be simple and performed using appropriate functions in `ggplot2`.
- Try and find a data set that you could link with the tibble and generate insights from this (as part of your three questions).
- Encode all your analysis in a quarto markdown file, and submit this along with a PDF output.



Dataset 1: nyc_bikes in the CRAN package tsibbledata

```
library(dplyr)
library(tsibbledata)

glimpse(nyc_bikes)
```

```
Rows: 4,268
Columns: 12
$ bike_id      <fct> 26301, 26301, 26301, 26301, 26301, 26301, 26301, ~
$ start_time   <dttm> 2018-02-26 19:11:03, 2018-02-27 07:52:49, 2018-02-27 12-
$ stop_time    <dttm> 2018-02-26 19:15:40, 2018-02-27 07:58:13, 2018-02-27 12-
$ start_station <fct> 3186, 3203, 3202, 3638, 3638, 3187, 3638, 3639, 3202, 36-
$ start_lat     <dbl> 40.71959, 40.72760, 40.72722, 40.72429, 40.72429, 40.721-
$ start_long    <dbl> -74.04312, -74.04425, -74.03376, -74.03548, -74.03548, --
$ end_station   <fct> 3203, 3202, 3638, 3638, 3187, 3638, 3639, 3202, 3638, 32-
$ end_lat       <dbl> 40.72760, 40.72722, 40.72429, 40.72429, 40.72112, 40.724-
$ end_long      <dbl> -74.04425, -74.03376, -74.03548, -74.03548, -74.03805, --
$ type          <fct> Subscriber, Subscriber, Subscriber, Subscriber, Subscrib-
$ birth_year    <dbl> 1986, 1979, 1963, 1984, 1980, 1980, 1989, 1962, 1992, 19-
$ gender         <fct> Male, Male, Male, Male, Male, Male, Male, Male, Male, Ma-
```

Dataset 2: flights in the CRAN package nycflights23

```
library(dplyr)
library(nycflights23)

glimpse(flights)
```

Rows: 435,352
Columns: 19

	Type	Description
\$ year	<int>	2023, 2023, 2023, 2023, 2023, 2023, 2023, 2023, 2023, 2~
\$ month	<int>	1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1~
\$ day	<int>	1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1~
\$ dep_time	<int>	1, 18, 31, 33, 36, 503, 520, 524, 537, 547, 549, 551, 5~
\$ sched_dep_time	<int>	2038, 2300, 2344, 2140, 2048, 500, 510, 530, 520, 545, ~
\$ dep_delay	<dbl>	203, 78, 47, 173, 228, 3, 10, -6, 17, 2, -10, -9, -7, --
\$ arr_time	<int>	328, 228, 500, 238, 223, 808, 948, 645, 926, 845, 905, ~
\$ sched_arr_time	<int>	3, 135, 426, 2352, 2252, 815, 949, 710, 818, 852, 901, ~
\$ arr_delay	<dbl>	205, 53, 34, 166, 211, -7, -1, -25, 68, -7, 4, -13, -14~
\$ carrier	<chr>	"UA", "DL", "B6", "B6", "UA", "B6", "AA", "UA", "A~
\$ flight	<int>	628, 393, 371, 1053, 219, 499, 996, 981, 206, 225, 800, ~
\$ tailnum	<chr>	"N25201", "N830DN", "N807JB", "N265JB", "N17730", "N925~
\$ origin	<chr>	"EWR", "JFK", "JFK", "JFK", "EWR", "EWR", "JFK", "EWR", ~
\$ dest	<chr>	"SMF", "ATL", "BQN", "CHS", "DTW", "MIA", "BQN", "ORD", ~
\$ air_time	<dbl>	367, 108, 190, 108, 80, 154, 192, 119, 258, 157, 164, 1~
\$ distance	<dbl>	2500, 760, 1576, 636, 488, 1085, 1576, 719, 1400, 1065, ~
\$ hour	<dbl>	20, 23, 23, 21, 20, 5, 5, 5, 5, 5, 6, 5, 6, 6, 6, 6, ~
\$ minute	<dbl>	38, 0, 44, 40, 48, 0, 10, 30, 20, 45, 59, 0, 59, 0, 0, ~
\$ time_hour	<dttm>	2023-01-01 20:00:00, 2023-01-01 23:00:00, 2023-01-01 2~