

# Fourth Industrial Revolution (4IR) Summer School

## Data Analysis Foundations – Day 3 exercises

---

### Question 1 [Daily Female Births]

In this Question, we will use the Daily Female Births dataset (Birth.csv). This time series dataset describes the number of daily female births in California in 1959. The units are a count and there are 365 observations.

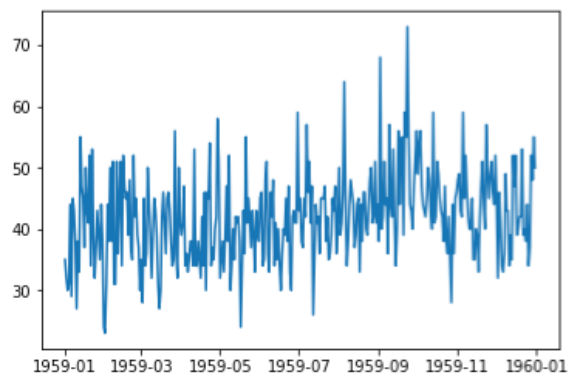
Date	Births
1959-01-01	35
1959-01-02	32
1959-01-03	30
1959-01-04	31

Now, do the following tasks:

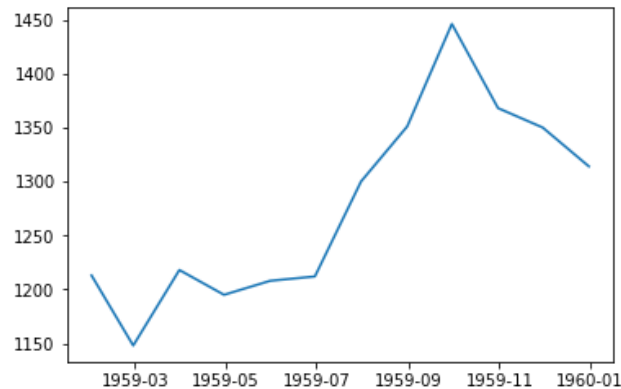
1. Load Time Series Data
2. Query by time to access all observations in January. What is the total number of female births in January?
  - (You should get 31 observations)
3. Print data descriptive statistics including mean, standard deviation, median, minimum, and maximum of the observations.

count	365.000000
mean	41.980822
std	7.348257
min	23.000000
25%	37.000000
50%	42.000000
75%	46.000000
max	73.000000

4. Plot time series data over the year



5. Resample the time series based on month. Then, plot it.



## Question 2 [Chicago Crimes] - Optional

Consider the Chicago crimes dataset hosted at Kaggle.

Either download it, or directly import it in your colab.

- Download the (Chicago\_Crimes\_2012\_to\_2017.csv) from kaggle.
- <https://www.kaggle.com/currie32/crimes-in-chicago>
- <https://towardsdatascience.com/setting-up-kaggle-in-google-colab-ebb281b61463>

Perform Exploratory data analysis to answer the following questions:

1. Which year has the most crimes?
  - Support your answer with line plot or a bar chart
2. For each year, which month has the highest number of crimes.
  - Support your answer with line plot or a bar chart
3. What is the highest type crime activity done form 2012 to 2017.
4. Can you detect any seasonal pattern?
  - Plot the data after resampling based on a year/month/week/day.

## Additional Exercises

### Question [Feature Selection]

KC1 dataset (KC1.csv) is a software fault dataset that is publicly available. It is used widely in the area of defect prediction using machine learning techniques. Choose and perform a feature selection technique, and investigate the selected features.

- Nine features (software metrics).
- One target (faulty). Binary feature.

### Question [Altair]

Open Altair jupyter notebook to find Altair plotting exercises.