



COURSE SYLLABUS

I- Course Information :

Title of the course	Data Visualization
Department	Data Science
Specialization	Data Science
Academic Year	2020 - 2021
Semester Duration	4 months
Credit Number	
Number of Hours	40 hours
Course Instructor	Mira Abboud
Email	mira.abboud@ul.edu.lb
Office Hours	2 hours / week

II- Course Description

One of the key skills of a data scientist is the ability to tell a compelling story, visualizing data and findings in an approachable and stimulating way. In this course, you will learn how to leverage a software tool to visualize data that will also enable you to extract information, better understand the data, and make more effective decisions. Mainly, the course includes: python basic libraries for data science (numpy, pandas) and all the techniques for basic, advanced, geospatial, maps, network graphs visualizations.

III- Course Outline and Calendar

Learning Objectives	Session(s)	Learning Content	Pedagogical Strategy& Tools
Python for data science	Session 1	Course syllabus + data science process + introduction to data visualization (how it fits in each step of the process)	Interactive course (chapter 0)
	Sessions 2-3	- Python basics: lists, tuples, sets, dictionaries - Numpy: basics, numpy array and main operations, read images, apply filters,	Interactive course (chapter 1) + Python implementation on Jupyter notebook + provide cheat sheet



		broadcasting, boolean ndarrays, arithmetic operations, compute stats, ... - Pandas (Series and DataFrames), read datasets, DataFrame multi-layers, indexing, filtering, slicing, descriptive stats, transpose, data Preprocessing (clean-up, standardization, normalization, scaling, transform, ...)	
Visualization basics	Session 4	Introduction to Matplotlib (its architecture composed of 3 layers + basic plotting) + understand Canada immigration dataset provided from IBM (which we will work on it during this course)	Interactive course (chapter 1) + Python implementation on Jupyter notebook
	Session 5	- Line plots: single plot, multiple plots, subplots, 3D plotting, annotate, styling(labels, background, grid,...) - Line plots for Time Series: date and time objects, resampling, up/downsampling data series, time zones handling, timedeltas, plotting	Interactive course (chapter 1) + Different Python Implementations : matplotlib Library (with scripting layer and the artist layer) - Seaborn Library implementation
		Assign graded homework on Times Series data. Deadline: to be delivered before session 8	
	Session 6	Area plots: single, multiple plotting, subplots, styling (labels, background,...), 3 types of plotting (overlapping, stacked, percentage), 3D plotting	Interactive course (chapter 1) + Different Python Implementations : matplotlib Library - Seaborn Library implementation
	Session 7	- Histograms: single, multiple plotting (side-by-side, stacked), analyze and transform distributions, 3D plotting, styling - Kernel Density Estimate (KDE): single, multiple plotting, probability density with histogram, shaded density plot, rug plots, 3D plotting, styling	Interactive course (chapter 1) + Different Python Implementations : matplotlib Library - Seaborn Library implementation



	Session 8	- Correct homework - Barcharts: single, multiple plotting, error bars, annotate, Horizontal Bar Plot, stacked plotting, 3D plotting, styling	Interactive correction + Interactive course (chapter 1) + Different Python Implementations : matplot Library - Seaborn Library implementation
	Assign projects on different topics. Deadline: to be delivered before session 22		
	Session 9	- Pie charts: group by, plotting, multiple plotting, subplots, styling, 3D plotting - Box plot and violin plot: plotting, multiple plotting, subplots, styling, 3D plotting	Interactive course (chapter 1) + Different Python Implementations : matplot Library - Seaborn Library implementation
	Session 10	- Scatter plot: plotting, multiple plotting, subplots, styling, 3D plotting - Bubble plot: plotting, multiple plotting, subplots, styling, 3D plotting - Lollipop plot: plotting, multiple plotting, subplots, styling	
Models visualization	Sessions 11-13	10 mns quiz + correction	
		+ General introduction to machine learning (1 simple algorithm will be explained for each category: regression, classification, clustering) + visualize (trained models, training and testing sets)	Interactive course (chapter 2) + Different Python Implementations : matplot Library - Seaborn Library implementation
Advanced Visualization Tools	Session 14	- Heat Map: plotting, multiple plotting, subplots, styling - Tree Map: plotting, multiple plotting, subplots, styling	Interactive course (chapter 3) + Different Python Implementations : matplot Library - Seaborn Library implementation
	Sessions 15-17	Waffle charts: introduction + Implement a full module for Waffle chart visualization (since no library for waffle charts in python)	Interactive course (chapter 3) + Implement a module for Waffle charts from scratch + check the PyWaffle which's under construction
	Session 18	Word Clouds: stopwords, interpolations, plotting + learn to generate text from the Canada dataset	Interactive course (chapter 3) + Python Implementation using word_cloud package
	Session 19	10 mns quiz + correction	



		Other plots (Venn diagram, radar plot, cat plot, swarm plot, strip plot): their use + examples + joint plots	Interactive course (chapter 3) + Python Implementation
Creating Maps and Visualizing Geospatial Data	Session 20	Introduction to Folium : create maps using longitude and latitude, zoom parameters + Map Styles (stamen terrain, stamen tonner, ...) + Maps with markers(create feature groups, children, marker function, clusters of markers)	Interactive course (chapter 4) + Python Implementation using Folium Library
	Session 21	Choropleth Maps: Geojson file, convert a map to a choropleth map	Interactive course (chapter 4) + Python Implementation using Folium Library
Network graphs	Session 22	Network graphs (Networkx)	Interactive course (chapter 5) + Python Implementation using Networkx, matplotlib and dash Libraries
Maths plots	Sessions 23-24	Polar plots (maths functions) + rendering text with latex	Interactive course (chapter 6) + Python Implementation using Matplot Library
	Sessions 25 - 26	Projects presentation + revision	Students have to present their projects
Final			

IV- Teaching approach & methods

- Most of the sessions will follow the Interactive approach
- Practice (implemented examples) will be a key factor
- Homeworks + research are included

V- Assessment Methods

The assessment of the student is carried out according to the following grid:

Assessment Type	Final Grade Weighting	%	Dates/Deadlines
Assignment	/30	40% of the final grade	Session 5 – session 8
10 mns quiz 1	/15		Session 11
10 mns quiz 2	/15		Session 19
Project	/40	60% of the final grade	Session 8 – session 22
Final exam	/100		Date: End of the course
Total		100%	

VI- Specific Course Requirements



1- Regulations concerning deadlines for written and applied work submissions

The teacher has the right to:

- Refuse to receive written applied work outside the deadlines.
- Refuse homework/project copies with high percentage of similarities.
- Refuse homework/project not submitted on MS teams.
- Give the student additional time after omitting up to 25% of the mark awarded to the concerned activity.

2- Academic measures in case of plagiarism

The student shall be liable to the following penalties in case of any attempt of illicit scientific citation:

- Cancellation of the student's contribution, activity or scientific article and reduction of the final mark.
- If the same offense is reported in other courses, cancellation of the final exam and failure in the relevant course(s).

VII- Reference Books

Milovanović, Igor. *Python data visualization cookbook*. Packt Publishing Ltd, 2013.

VIII- Bibliography & Additional Sources

Embarak, Ossama. "Data analysis and visualization using python.", 2018