

Exploratory Data Analysis

Agenda

Activities	Objective	Duration	Time
Introduction	<ul style="list-style-type: none"> - Quick Check in - Google Colabs 	3 mins	12:00
Data Transformation	<ul style="list-style-type: none"> - Clean and rearrange data - Slice and merge data 	5 mins talk 15 mins group work	12:05
Data Visualization	<ul style="list-style-type: none"> - Basic Plots - Themes + color - Interactive plots 	5 mins talk 20 mins group work	12:30
Break	- Chat + Questions	15 mins	1:05
Statistics	<ul style="list-style-type: none"> - Time Series - Correlation - Variation + Bias 	10 mins talk 20 mins group work	1:20
Modeling	<ul style="list-style-type: none"> - Mathematical - Statistical 	5 mins talk 25 mins group work	1:55
Break	- Chat + Questions	5mins	2:30
Presentation	<ul style="list-style-type: none"> - Prepare 3 Slides - Present 	15 mins prep 60 mins (6mins/grp)	2:35
Closing	- Remarks	5mins	3:55

NB: After every group work there is 5 mins extra to be used however you desire.

Why group work?

‘Teamwork makes dream work’

As an AI tech person, you will always be required to work with people. In a world where norms are changing, remote work has become the new normal. Therefore, gaining experience in working with people from different background and time zone is what this session is offering.

I know you can complete these exercises alone, but I want you to do it in a group. Take this as a time to showcase your excellent team skills and to learn from each other.

Come lets Bootcamp together!!!

Prep materials

These links will help you prepare for the group work. The knowledge would be required for your group to complete its challenge without stress.

1. Colabs

This is similar to jupyter notebook by Google. It will be important for the group collaboration.

2. Data Transformation

Sometimes the data given to us cannot be used in its original format, hence, the need for transformation. You will focus on how to clean, slice, and merge data.

<https://github.com/Ebude/Data-Wrangling-with-Python> play with data.

3. Visualization

This section is to enable you gain knowledge on quick and effective visualization tools. You will learn to plot a graph as well as interpret the graph. These are the packages required:

- **Matplotlib:** <https://matplotlib.org/>
- **Seaborn:** <https://seaborn.pydata.org/>
- **Plotly Express:** <https://plotly.com/python/plotly-express/>

4. Statistics

The notion of basic statistics if necessary for understanding the data before going on to select the right AI model. You will carry out a statistic test on the data, here are a few notions required (you don't need a mastery level, just the basic is sufficient):

- **Stationality:** <https://www.analyticsvidhya.com/blog/2018/09/non-stationary-time-series-python/>
- **Seasonality:** <https://machinelearningmastery.com/decompose-time-series-data-trend-seasonality/> decompose the data into trends, seasonal, and error
- **Variance:** <https://appdividend.com/2019/08/08/python-variance-example-python-statistics-variance-function/>
- **Correlation:** <https://machinelearningmastery.com/how-to-use-correlation-to-understand-the-relationship-between-variables/> introduction and spearman's correlation

- **Autocorrelation:** <https://www.machinelearningplus.com/time-series/time-series-analysis-python/> go to number 16 only

5. Modelling

This is a quick prescriptive or predictive analysis on data in order to select the right AI model. In this section, you will focus on these two class: Statistical and Mathematical models. Here are a few ideas:

- **SARIMA:** <https://towardsdatascience.com/time-series-in-python-part-2-dealing-with-seasonal-data-397a65b74051>
- **One-way ANOVA:** <https://reneshbedre.github.io/blog/anova.html>

6. Presentation

Most times data scientists are required to present their findings to the company before proceeding with ML modeling. Presentation skills of data analysis comes in very handy. Your group is required to explain to the audience what was discovered in the data and which ML model would suit the data and the objective. This would be done in a 5 mins pitching style.