**Data Analysis with python**

**Tentative plan (9pm Pak time Monday to Friday, 29th Nov. to 28th Jan.)**

Prerequisite: introduction to python language

***Contents:***

**Chapter-1: Data Manipulation with pandas**

1. Let’s master the pandas basics. Learn how to inspect DataFrames and perform fundamental manipulations, including sorting rows, subsetting, and adding new columns.
2. you’ll calculate summary statistics on DataFrame columns, and master grouped summary statistics and pivot tables.
3. Indexes are supercharged row and column names. Learn how they can be combined with slicing for powerful DataFrame subsetting.
4. Learn to visualize the contents of your DataFrames, handle missing data values, and import data from and export data to CSV files.

**Chapter-2: Merging DataFrames with pandas**

1. you'll learn about different techniques you can use to import multiple files into DataFrames. Having imported your data into individual DataFrames, you'll then learn how to share information between DataFrames using their indexes.
2. you'll learn about appending and concatenating DataFrames
3. You'll explore different techniques for merging, and learn about left joins, right joins, inner joins, and outer joins, as well as when to use which. You'll also learn about ordered merging, which is useful when you want to merge DataFrames with columns that have natural orderings, like date-time columns.

**Chapter-3: Introduction to Importing Data in Python**

1. you'll learn how to import data into Python from all types of flat files, which are a simple and prevalent form of data storage.
2. you'll learn how to import data into Python from a wide array of important file types. These include pickled files, Excel spreadsheets, SAS and Stata files, HDF5 files, a file type for storing large quantities of numerical data.
3. You will learn about relational models, how to create SQL queries, how to filter and order your SQL records, and how to perform advanced queries by joining database tables.

**Chapter-4: Intermediate Importing Data in Python**

1. you will learn how to get data from the web, whether it is stored in files or in HTML. You'll also learn the basics of scraping and parsing web data.
2. You will learn the basics of extracting data from APIs, gain insight on the importance of APIs, and practice extracting data by diving into the OMDB and Library of Congress APIs.
3. you will consolidate your knowledge of interacting with APIs in a deep dive into the Twitter streaming API. You'll learn how to stream real-time Twitter data, and how to analyze and visualize it.

**Chapter-5: Cleaning Data in Python**

1. you'll learn how to overcome some of the most common dirty data problems. You'll convert data types, apply range constraints to remove future data points, and remove duplicated data points to avoid double-counting.
2. you’ll learn how to fix whitespace and capitalization inconsistencies in category labels, collapse multiple categories into one, and reformat strings for consistency.
3. advanced data cleaning problems, such as ensuring that weights are all written in kilograms instead of pounds.
4. Record linkage is a powerful technique used to merge multiple datasets together, used when values have typos or different spellings.

**Chapter-6: Exploratory Data Analysis in Python**

1. The first step of almost any data project is to read the data, check for errors and special cases, and prepare data for analysis.
2. you'll learn how to represent distributions using Probability Mass Functions (PMFs) and Cumulative Distribution Functions (CDFs)
3. you'll explore relationships between variables two at a time, using scatter plots and other visualizations to extract insights from a new dataset
4. Explore multivariate relationships using multiple regression to describe non-linear relationships and logistic regression to explain and predict binary variables.

**Chapter-7: Probability**

1. The Basic Probability Formula, Computing Expected Values, Frequency, Events and Their Complements.
2. Fundamentals of Combinations, Permutations and How to Use Them, Simple Operations with Factorials, Solving Variations with Repetition, Symmetry of Combinations.
3. Sets and Events, union and Mutually Exclusive sets, Dependence and Independence of Sets, The Conditional Probability Formula, The Additive Rule, multiplication law, Bayes' Law.
4. Probability Distributions; uniform, Bernoulli, Binomial, poisson, Continuous, normal, standard normal, Exponential, logistic etc

**Chapter-8: Statistical Thinking in Python**

1. first explore your data by plotting them and computing simple summary statistics
2. Levels of Measurement, Categorical Variables - Visualization Techniques, Numerical Variables - Frequency Distribution Table, Histogram,
3. Cross Tables and Scatter Plots, Mean, Median and Mode, Skewness, Variance, Standard Deviation, Covariance, Correlation.