

# Mentorskool

What learning path suits me?



# Questions to ask yourself



#### Where should I start?

## Possible responses:

- I am a fresher
- I am experienced person but new to Data Analytics
- I use Excel in my day to day work
- I have started off with SQL
- I know SQL and started off with visualization in tools like Tableau / PowerBI
- I started working on predictive modelling/machine learning

## Should I aim for covering breadth or depth?

# Possible responses :

- I don't know, you help me
- I like math. I think going deep in ML/AI is my cup of tea
- I am a creative person. Visuals are my forte
- o I like working with messy data. SQL works for me



# Data Science Skills Level - I



Problem Analysis

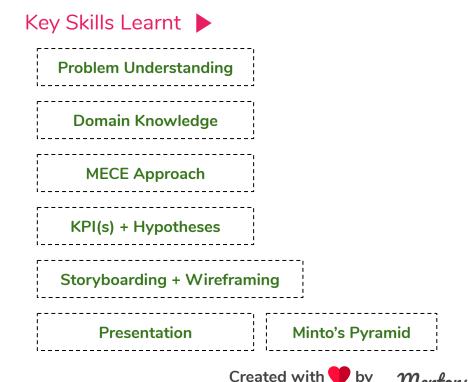
**KPI Building** 

Hypothesis Building

Storyboarding

# Learning Topics >

- Understanding the problem
- Identifying the industry domain
- Learning the MECE Approach
- Developing Hypotheses
- Developing KPIs
- Storyboarding the solution
- Developing wireframes to visualize the solution
- Build story flows on ppt/google slides/jupyter notebooks
- Learning the right communication approach - Minto's Pyramid



# Data Science Skills Level - II



- Ability to query data from a variety of sources
  - Relational databases : MySQL, MS-SQL Server, Postgresql
  - NoSQL databases : MongoDB, Cassandra, DynamoDB etc.
  - Hadoop Distributed File System
  - Web API
  - Excel, Text, CSV, JSON, XML etc.

## Ability to pre-process data

- Knowledge of simple to complex SQL queries
- Brief idea of SQL functions and procedures
- Extensive knowledge of data pre-processing in pandas
- Missing value treatment
- Outlier Treatment

## Ability to build simple to complex reports and visualizations

- Knowledge of at least one BI & Visualization tools like Tableau/Power BI
- Knowledge of data warehouse concepts ex. Facts & Dimensions
- Visualizing data using Python (Matplotlib, Seaborn, Plotly etc.)



# Data Science Skills Level - III



#### Ability to perform Statistical analysis on data

- Ability to summarize data using statistical properties
- Ability to perform exhaustive univariate analysis.
- Idea of statistical tests
- Ability to perform bivariate/multivariate analysis
- Understanding of hypothesis testing
- Good grasp over the idea of normal distribution
- Excellent skills at Exploratory Data Analysis

### Ability to explain inner workings of Machine Learning models

- Knowledge of types of ML problems (Supervised, Unsupervised)
- o Idea of the mathematics of ML models
- Knowing how to handle different types of data
- Knowing the steps to implement before the data is ready to be fed into ML models
- Ability to explain the output of models
- Ability to explain assumptions which are required for a model
- Ability to tune the model to arrive at the best results
- Ability to explain why a model should work in a given situation?





# Data Science Skills Level - IV



#### Neural Networks

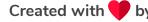
- Brief idea of Multi-layer perceptrons
- Brief idea of inner workings of a neural network. Understanding of backpropagation, activation functions, optimizers etc.
- An understanding of when neural networks is NOT to be used

### Computer Vision

- Knowledge of Image processing techniques
- Brief idea of Convolutional neural networks
- An idea of face detection and recognition techniques
- Decent experience in implementation of CNNs both by scratch as well as by transfer learning
- o Good knowledge of computer vision libraries like PIL, OpenCV, dlib etc.

### Natural Language Processing

- Knowledge of text processing techniques
- Understanding traditional NLP algorithms
- Learning NLP the deep learning way: CBOW, Skip-grams
- Learning about common Python packages for NLP: nltk, gensim, fasttext etc.





# Data Science Skills Level - V



#### Working with Cloud

- Ability to perform all the tasks mentioned earlier on a Cloud platform (Azure, AWS, GCP)
- Ability to design a data science pipeline using Cloud components
- Deploy an end-to-end data science project in Production on Cloud

### Technologies Required

- SQL for data mining
- Python for everything!
- Spark for data mining at Scale!
- Power BI for visualization