**INTRO | DEFINE**

During any activity on internet you are producing tunes of data, by using the data to make the decision is called data science. Guido van Rossum, 1980 Dutch Programmer Netherland. **1-M**

**USEAGE**

1. Data analysis
2. AI & ML
3. Automation
4. Web, Desktop & Mobile
5. Software Testing
6. Hacking
7. Bioinformatics (Vaccine)
8. Research
9. Big data, Meta data
10. Deep Learning | Predictive Modeling

**USERS**

Scientist, Data Analyst, Math, Stat, Accounts, Network, Sundas Khalid, faang Company

**STEPS TO LEARN**

1. Basic Math
2. Statistics
3. Python | R
4. Data Wrangling & Data Cleaning
5. Data Visualization
6. Machine / Deep learning Algorithms
7. Linux/ Conda / Git Environments
8. Open source Database / GitHub / Kaggle.com
9. Domain specific data science projects
10. Motivation / Good PC

**BENEFITS**

**GOAL / AIM /TARGET REMOTE JOB**

**WORKING FLOW CHART/ ROUTINE**

**01-01-23 40/8= 5 Week**

**Week 1 Python Programming**

**Week 2 Basic Math, Statistics, Inter mediate Statistics, Statistics+ Data Visualization, logic building, Data Driven Decisions, project to publication ready graph (Start to end)**

**Week 3 projects (Collect Data, manage data, tidy up data, and clean data, missing values, normal distribution) (write in note book & publish on kaggle, GitHub & tweeter)**

**Week 4 Machine learning, Image Analysis, prediction model**

**Week 5 Case Studies, projects, paper publication (See)**

**Week 6 Week 7 Week 8**

**Commands in CMD**

* **Python –version**
* **Python paste file in cmd**
* **Dir**
* **Cd desktop**
* **mkdir folder name**
* **cd .. (Back)**
* **cd . (Same)**
* **code .**

**Python Extension**

* **IntelliCode**
* **Python**
* **Top 5 jupyter notebook**
* **Codesnap**
* **Mark down all in one**
* **Prettier**
* **Blackbox**

**File Types**

* **.py**
* **.ipynb (Notebook)**

**How to use mobile camera as web cam for PC**

* **Iriun Application**

**GitHub Education Pack Resources**

* **GitHub Campus Expert**

**Rules**

* **DEMAS**
* **PEMDAS**
* **BDMAS**
* **DMAS, BODMAS or PEMDAS**

**# Assignment: What is type casting and how do we use that in python**

**# Assignment: when to use comments in python? Mention 10 study cases**

**Data**

**Data = insights for our future work**

**To exclude hidden secrets is known as data science**

**To exclude science behind the data**

**Numeric & Strings**

**Column, Row & unit, Variables & Indexes**

**Header = Column/ Row heading**

**Height Feet/CM/Inchs**

**Weight KG**

**Data Collection/ Generation**

**Original Data**

**Primary 1st used**

**Secondary 2nd used**

1. **Kaggle.com**
2. **Google data search**
3. **github.com**
4. **faostats.com**
5. **Websites (data Scraping, Collect data from any website and take prediction)**
6. **Stats**
7. **NADRA**
8. **Crime Detection**
9. **DNA (fingerprint detection further forensics)**
10. **Business analytics (Stock Exchange)**
11. **University database**
12. **Research database (Meta data analysis, generate new data by using previous data)**

**Python**

**Data**

1. **Visualization**
2. **Analysis (EDA)**
3. **Reports**
4. **ML DL**

**Plotting**

**Scatter plot / point plot**

**Bar plot**

**Box plot**

**Stack plot**

**Line plot**

**Histogram plot**

**Normality Normal distribution**

**Maps plot**

**Stacked plot**

**Pie plot (%age)**

**Pair plot**

**Data points**

**Canvas**

**Axis**

**Text**

**Information**

**Annotation**

**EDA Exploratory Data Analysis**

**Sense / Explore**

**ML**

**Data**

**Training**

**Model**

**Final Output**

**Prediction**

**Train Machines**

**Recognize Patterns and suggest / predict**

**e. g. Google Maps**

**Decision making on the base of Data**

**Prediction**

**4 – Types**

**Supervised mostly used (Data, Labels, Values are given) (x, y) (x, ?) (? , y)**

**Regression -> Number, Quantity**

**Classification -> Category, Alphabet**

**Un-Supervised**

**Make Clusters on Category basis**

**Semi Supervised-> partially labeled**

**Reinforcement e. g. ChatGPT (Learn from Mistakes)**

**Train Test Split**

**Training Data**

**Training 75% to 80%**

**Testing Data**

**20% to 30%**

**Day\_17**

**Data**

**Learning Prediction**

**Decisions**

**Simple linear regression**

**Y=a+ b(x)**

**A = intercept / Constant**

**B = Slope of function**

**Multiple linear regression**