All Rounder Preparation

This is a basic data Science roadmap with links and instructions as to how to become a proper data scientist. It may inlude necessary and sometimes not so neccessary but helpful topics to work upon. Use it to your own personal discretion and keep it updated when necessary.

Serial Number	Topic Name	Platform 1	Platform 2	Date Completed			Personal Difficulty 1-5			Notes		
				Hemant Gamin	Prateek Dilaware	Tanmay Somani	Hemant Gamin	Prateek Dilaware	Tanmay Somani	Hemant Gamin	Prateek Dilaware	Tanmay Somani
1	**Fundamentals**											
1 1	Matrices and Linear Algebra	Linear Algebra and Matrices				22-01-24			1			Started, to be continued
1 2	Hash Function, Binary Tree ,O(n)	Hash Functions , Binary Tree	www.geeksforgeeks.org/introduction-to-hashing-data-structure-and-algorithm-tutorials/		09-02-2024	23-01-24		3	3		**Binary search tree remains**	LinkedIn Lists is a prerequisite, binary tree again
1 3	Relational Algebra , DB Basics	Relational Algebra GATE Notes, Db for Beginners				25-01-24			1			very basic though must remember the notations
1 4	Inter, Outer, Cross, Theta Join	Joins				25-01-24			2			easy enough, just remember cross join
1 5	CAP Theorem	Basics of CAP				26-01-24			2			ACID is different.
1 6	Tabular Data					26-01-24			1			
1 7	Entropy	Entropy							2			
1 8	Data Frames and Series	DataFrame and Series				29-01-24			2			
1 9	Sharding	Db sharding				30-01-24			1			also see replication
1 10	OLAP	OLAP				31-01-24			1			also read OLTP
1 11	Multidimensional Data Model	Multidimensional Data Model				31-01-24			1			
1 12	ETL					01-02-24			1			
1 13	Reporting Vs BI Vs Analytics	Reporting Vs BI , BI Vs Big				01-02-24			2			easy definition and fundamental differences
1 14	JSON & XML	JSON VS XML , Json Vs Xml				01-02-24			1			
1 15	NOSql	NoSQL				01-02-24			2			need deeper understanding
1 16	Regex	Regex							2			
1 17	Vendor Landscape											cannot understand at all
1 18	Env. Setup	IDE_VCS_Gradle, Maven, or Webpack_Libraries,D	_			01-02-24			2			Personal Choices
2	**Statistics**											
2 1	UCI repository	UCI										Need to start working on datasets for now onto stat
2 2	Descriptive Statiscstics (mean,median,mode,range, SD)	Descriptive Statistics , Descriptive Statistics				02-02-24			1			clearly understood
2 3	Exploratory Data Analysis	EDA	EDA									do it again.
2 4	Histograms											
2 5	Permities & Outliers											
2 6	Probability Theory	Probability Theory				04-02-24			2			
2 7	Bayes Theorem	Bayes Theorem										
2 8	Random Variables											
2 9	Cumulative Distribution Function											
2 10	Continuous Distributions											
2 11	Skewness											
2 12	ANOVA											
2 13	Probability Density Function											
2 14	Central Limit Theorem											
2 15	Monte Carlo Method											
2 16	Hypothesis Testing											
2 17	P value											
2 18	Chi ^ Test											
2 19	Estimation											
2 20	Confidence Interval											
2 21	MLE											
2 22	Kernel Density Estimate											
2 23	Regression											
2 24	Covariance											
2 25	Correlation											
2 26	Pearson's Coefficient											
2 27	Causation											
2 28	Least-Fit											
2 29	Eucleadean Distance											
3	**Programming**											
3 1	Python Basics											
3 2	Working in Excel											
3 3	R Studio Setup											
3 4	Expressions											
3 5	IBM SPSS											
3 6	Variables											
3 7	Rapid Miner											
3 8	Vectors											
3 9	Matrices											
3 10	Arrays											
3 11	Factors											
3 12	Lists											
3 13	Data Frames											
3 14	Reading CSV/Raw Data											
3 15	Subsetting Data											
3 16	Manipulate Data Frames											
3 17	OpenCV	https://www.youtube.com/watch?v=oXlwWbU8l2o										
3 18	Functions											
3 19	Factor Analysis											
3 20	Install Packages											
3 21	Exploratory Data Analysis											
4	**Machine Learning**											
4 1	What is Machine Learning?											
4 2	Numerical Variance											
4 3	Categorical Variance											
4 4	Supervised Learning											
4 5	Unsuperized Learning											
4 6	Concepts, Inputs and Attributes											
4 7	Training, Input and Attributes											
4 8	Training & Test Data											
4 9	Classifier											
4 10	Prediction											
4 11	Lift											
4 12	Overfitting											
4 13	Bias and Variance											
4 14	Trees and Classification											
4 15	Classification Rate											
4 16	Decision Trees											
4 17	Boosting											
4 18	Naive Bayes Classifiers											
4 19	K-N-N											
4 20	Logistic Regression											
4 21	Ranking											
4 22	Linear Regression											
4 23	Perceptron											
4 24	Clustering											
4 25	Neural Networks											
4 26	Sentimental Analysis											
4 27	Collaborative Filetring											
4 28	Tagging											
5	**Text Mining / NLP**											
5 1	Corpus											
5 2	Entity Recognition											
5 3	Text Analysis											
5 4	UIMA											
5 5	Term Document Matrix											
5 6	Term Frequency And Weight											
5 7	Support Vector Machines											
5 8	Association Rules											
5 9	Market Based Analysis											
5 10	Feature Extraction											
5 11	Using Mahout											
5 12	Using Weka											
5 13	Using NLTK											
5 14	Classify Text											
5 15	Vocabulary Mapping											
5 16	Tagging											
6	**Visualization**											
6 1	Data Exploration in R											
6 2	UniBi and Multi Variate Viz											
6 3	ggplotz											
6 4	Histogram and Pie											
6 5	Tree and Tree Map											
6 6	Scatter Plot											
6 7	Line Charts											
6 8	Spatial Charts											
6 9	Survey Plot											
6 10	Timeline											
6 11	Decision Trees											
6 12	D3.js	D3.js										
6 13	InfoVis											
6 14	IBM ManyEyes											
6 15	Tableu											
7	**Big Data**											
7 1	MapReduce Fundamentals											
7 2	Hadoop Components											
7 3	HDFS											
7 4	Data Replication principles											
7 5	Setup Hadoop											

7 6	Name and Data Nodes											
7 7	Job and Task Tracker											
7 8	MIR Programming											
7 9	Sqoop : Loading Data in HDFS											
7 10	Flume Scribe for unstructed data											
7 11	SQl with pig											
7 12	DWH with Hive											
7 13	Scribe chukwa for Weblog											
7 14	Using Mahout											
7 15	Zookeeper Avro											
7 16	Storm Hadoop											
7 17	Rhaddop Rhipe											
7 18	rmr											
7 19	Cassandra											
7 20	MongoDB,neo4j											
8	Data Ingestion											
8 1	Summary of Data Formats											
8 2	Data Discovery											
8 3	Data Sources and Acquisition											
8 4	Data Integration											
8 5	Data Fusion											
8 6	Transformation and Enrichment											
8 7	Principal Component Analysis											
8 8	Data Survey											
8 9	Google OpenRefine											
8 10	how much Data?											
8 11	Using ETL											
9	**Data Munging**											
9 1	Dimensionalities and Normalities Reduction											
9 2	Normalization											
9 3	Data Scrubbing											
9 4	Handling Missing Values											
9 5	Unbiased Estimators											
9 6	Bining Sparse Values											
9 7	Feature Extraction											
9 8	Denoising											
9 9	Sampling											
9 10	Stratified Sampling											
9 11	Principal Component Analysis											
10	**Toolbox**											
10 1	MS Excel w/ Data Analytics ToolPak											
10 2	Java,Python											
10 3	R, RStudio,Rattle											
10 4	Weka,Knime,Rapid Miner											
10 5	Hadoop											
10 6	Spark											
10 7	Storm											
10 8	Flume,Scibe,Chukwa											
10 9	Nutch,Talend,Scraperwiki											
10 10	Webscraper,Flume,Sqoop											
10 11	tm,Rweka,NLTK											
10 12	RHIPE											
10 13	d3.js,ggplot2,Shiny											
10 14	IBM language ware											
10 15	Cassandra,MongoDB											
10 16	SQL,PowerBI,Tableu											

S . No	Name	Link	Note
1	Learn Dancing	https://www.youtube.com/playlist?list=PLGN5lnbZpwnnLWEHfMMqdKOBimH1aXOZp	
2	Basic Data Science Pathway	https://t.me/machinelearning_course	
3	Aptitude Test Practise	https://www.javatpoint.com/aptitude/quantitative	
4	Machine learing guide		
5	Leetcode DSA	Leetcode Blind 75	
6	Foage Virtual Internships	https://www.theforage.com/	
7			
8			
9			

S.No.	Project Name	Links	Note
1	AI image classification		
2	Intelligent Traffic Management System and Challan		
3	Attendance Image classification		
4			
5			
6			
7			
8			
9			
10			

3 newsletters subscribed, Learn GitHub		Concepts of SQL	Documentation 2 [85%], 2 Presentations, Story using 50 new learnt words, 1 research ideas	Concepts of Python (Basic, Data Structures , Libraries, Pandas, numpy, matplotlib, seaborn, tkinter, pytorch or scikitlearn) [90%] ; pip , Project 3 ; WPM expected to be near 35, 60 percent score using careerflow	**Resume Creation**		Concept of Exc	**All resources Completed!!!, 1 Internship**		
3 Project Ideas, Learn GitHub	1 Project, 2 newsletter subscribed ,	Concepts of SQL	1 Project, Story using 50 new learnt words	1 Project, 1 SQL Project , WPM expected to be near 35, Concepts of Python(Basic, Data Sctructures). 60 percent using careerflow	**Resume Creation**	5 Speeches , 2 :3 ; 5:2	Concept of Excel	**All resources Completed!!!, 1 Internship**		
Linkedin Post, 7 Instagram posts, 1 Minor projects	Linkedin Post, 7 Instagram posts	Concepts of SQL,Linkedin Post,7 Instagram posts	Story using 50 new learnt words, Linkedin Post, 7 instagram posts , 1 Minor project	WPM expected to be near 50, Linkedin Post, 7 instagram post	**Resume Creation,** Instagram post	Linkedin Post, 7 Instagram post	Linkedin Post, 7 Instagram post, 1 Minor Project	**All resources Completed!!!, 1 Internship,** 1 Major Project	7 Instagram post	7 Instagram post

INTERNSHIP									
S No.	Date	Name of the company	Company website	Company Linkedin	Source	First mail/ first message	Follow up 1	Follow up 2	Conclusion

INTERNSHIP									
S No.	Date	Name of the com	Company websit	Company Linked	Source	First mail/ first m	Follow up 1	Follow up 2	Conclusion

INTERNSHIP									
S No.	Date	Name of the com	Company websit	Company Linked	Source	First mail/ first m	Follow up 1	Follow up 2	Conclusion