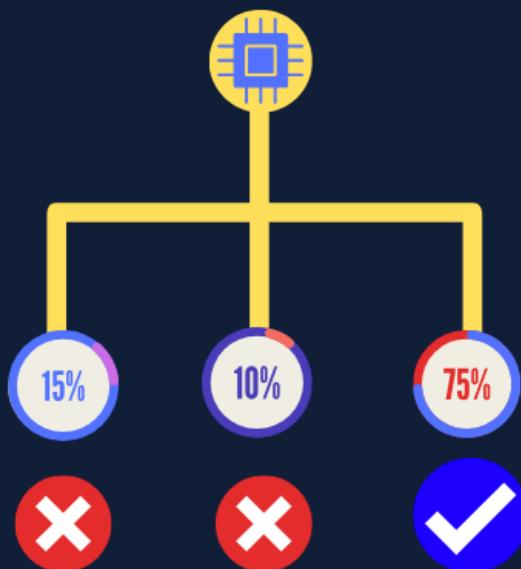


# Data Science & Machine Learning Developer track





# Key Focus: 9 skills

- Defining Data Science Problems
- Data Extraction using SQL & Web Scraping
- Data Cleaning Strategies
- Statistics for Data Science
- Exploratory Data Analysis
- Developing Data Pipelines
- Machine Learning
- Communicating Data Driven Insights
- Developing Data Science products
- Kaggle Master class
- Career Development Program



**alteryx**

# Target

- **Team Leaders & CXO's** who want to ace the data driven decision-making process and develop strategies for Business transformations into the A.I. world
- **Project managers** who aim to manage scalable data science projects
- **Product Managers** who aim to develop scalable Machine Learning Products
- **Software engineers & Developers** who want to code and develop application in Data Science Software and services
- **Software Testing engineers & ETL + Database developers** who want to upskill in Data Science and Machine Learning
- **Marketing, Sales, Business Development, Operations & Human Resources Managers** who want to apply Data Science & Machine Learning skills on real-world day to day business problems
- **SME & Domain experts** in Telecom, Retail, Finance, Healthcare, Banking, Oil&Gas, Smart energy, IOT, 3D Printing, Process Automation areas developing Data Science & Machine Learning Processes
- **Platform Experts** in SAP, Oracle BI, Salesforce, Teradata, who want to implement & develop advanced analytics solutions on their systems
- **Business & Strategy Consultants** who want to transition into data science consulting
- **Educators** who want to teach Machine Learning & Data Science
- **Students & Recent graduates** who want to start internships in Data Science & Machine Learning field
- **For all others** who want to upskill and become competitive in todays job market



# Data Science with Tableau



Learning Content from the perspective of a Product Manager & Data  
**Engineer @ Accenture & EdTex**

## Salient features of the Course

**Course Workload** ~ 120 hours

**Videos duration** ~ 15 hours duration (on demand)

**Product Development with Tableau session:** 2 hours duration (online)

**Learner Level:** Beginners to Advanced Users

**Tableau Desktop Certification:** Yes, provided by DataSciencePrep.Co

**Interview Prep for Tableau with 150 real interview Q&A:** Provided with A.I. Track

**More than 1000 IT & Business Professionals took this course since the year 2016**

**Total Companies** they represent > 50

**Total Countries** they represent > 10

Modules	Topics	Total Exercises + Case studies
1	Tableau Ecosystem	2
2	Data Visualization Fundamentals	2
3	Segmenting and Filtering data	3
4	Calculation fields	1
5	Functions in Tableau	1
6	Descriptive Statistics using Tableau	1
7	Data Joining and Blending in Tableau	1
8	Machine Learning with Tableau, R and Python	1
9	Tableau Projects with Certification Exam Material	5



# Data Science with Tableau

Become a Tableau Jedi

## Module 1: Introduction to Tableau Ecosystem

- Introduction to Visual Analytics
- Introduction to Tableau
- Introduction to Tableau Products (Desktop, Public and Online versions)
- Installing Tableau
- Loading **Excel, CSV, SQL data, JSON** into Tableau
- Live data connections vs data Extracts
- Introduction to Data source filters
- Inspection of data and editing metadata
- Categorizing Variables or Fields into “Dimensions Roles and Measures Roles”
- Introduction to Dates and Time fields
- Introduction to file types: *.tmb, .twb, .tde, .twbx, .tds, .tdsx, .tps, .tms*
- Introduction Tableau .hyper format
- Using Rows and Columns Shelves
- Saving Tableau workbooks(**.twb, .twbx formats, Tableau Public domain**)
- Saving Tableau visualizations into images and pdf formats
- **Hands-on Exercise One** - Creating first visualizations in Tableau on Retail Sales data
- **Hands-on Case Study** - Highest grossing Hollywood movie analysis from the Year 2000

## Module 2: Data Visualisation Fundamentals

- Introduction to data types
- Introduction to levels of measurements
- Discrete vs Continuous values
- Introduction to the Bar Graph
- Introduction to Line Graphs
- Introduction to Pie Charts
- Introduction to Tree Graphs
- Introduction to Histograms and scatter plot
- Introduction to Boxplots
- Introduction to Heat Maps
- Introduction to Bullet Graphs
- Introduction to Histograms

- Introduction to Gantt Charts
- Introduction to Data Tables
- Introduction to Bubble charts
- Sorting data
- Clearing sheets
- Introduction to Geographical Maps
- Custom Geocoding
- Connecting external map services
- Introduction to Dashboards
- Developing Dashboard structure
- Developing interactive dashboards
- When to use which visualizations?
- Formatting Visualizations
- What are the most effective Visualizations in Dashboards and Data Science reports?
- **Hands-on case study:** EarthQuake and GeoSpatial Data Analysis from the Year 1900 – 2013
- **Hands-on Case study:** Exploratory Data Analysis on sales data of Rossman Health Chain

### **Module 3: Segmenting and Filtering Data**

- Introduction to Tableau Pages Shelf, Filters shelf, Marks Card
- Introduction to Data Extract Filters
- Introduction to Data Source Filters
- Introduction to Context Filters
- Filters on Dimensions
- Filters on Measures
- Grouping Data
- Working with Static Sets
- Creating bins with continuous data
- Applications of Dynamic Sets
- Applications of Combined Sets
- Parameters controls
- **Hands-on Exercise two:** Top 10 Startups: Investment analysis
- **Hands-on Exercise three:** Applications of Parameter control on Rossman Health Sales data
- **Hands-on Exercise Four:** Application of parameters on Bank Customer segmentation

## Module 4: Calculation fields

- Introduction to Order of Operations in Tableau
- Data aggregations in Tableau
- Introduction to Calculation fields
- Types of Calculations in Tableau
- Row-level Calculations
- Table Calculations
- LOD calculations (INCLUDE, EXCLUDE, FIXED calculations)
- Transform calculations
- AdHoc calculations
- Best Practices for writing calculations in Tableau
- **Hands-on Exercise Five:** Retail sales analysis use LOD calculations

## Module 5: Data Joining and Blending in Tableau

- Introduction to SQL Joins
- Blending Data from CSV and SQL database formats
- Calculations on Blended data
- Editing and defining relations between categorical variables and date variables
- Dual Axis charts and applications
- **Hands-on Exercise Seven:** Retail Sales data analysis using data blending

## Module 6: Descriptive Statistics in Tableau

- Calculating the Mean
- Calculating the Median
- Calculating Percentiles
- Calculating the Standard deviation
- Calculating the variance
- Applications of Histograms
- Applications of Trendlines
- **Hands-on Case Study:** H1B Visa data analysis of Data Scientists from 2008 to 2015

## Module 7: Functions in Tableau

- Number Functions
- String Functions
- Date Functions
- Type Conversion

- Logical Functions
- Aggregate Functions
- **Hands-on Exercise Six:** Retail data analysis using inbuilt Functions

## **Module 8: Developing Machine Learning Products with Tableau & R**

- Understanding the Logic Behind Machine Learning
- Identifying the Predictor and Response variables
- Choosing Machine learning algorithms for Continuous and Class outcomes
- Applications of Simple Linear Regression
- Applications of Multiple Linear Regression
- Applications of Logistic Regression
- Interpreting Machine Learning Model Outcomes
- Improving Model performance at the basic level
- Forecasting Sales Tableau
- Clustering data Tableau
- Applying Trends lines in Demand forecasting scenarios
- Integration of R programming tool with Tableau
- Integration of Python tool with Tableau
- **Hands-on Exercise Eight:** Developing Linear Regression Dashboard Products for Automobile Performance Prediction
- **Hands-on Exercise Nine:** Developing Classification Dashboard Products for Cancer Detection
- **Hands-on Exercise Ten:** Sales Forecasting in Rossmann Healthcare (A Kaggle Exercise)

## **Module 9: Tableau – EDA Projects + Desktop 10 Certification Exam Material**

- Project 1:** Hubway data visualization challenge on Bike demand  
**Project 2:** Uber Taxi demand analysis  
**Project 3:** EDA of Walmart establishments growth from years 1970 to 2006  
**Project 4:** Wimbledon Championship Analysis from years 1877 to 2016  
**Project 5:** EDA of Airbnb listings in Newyork  
**Certification Exam Material for Tableau Desktop 10**

## **Module 10: Industry and Technical series from the Vaults of Tableau**

- Tableau Product Series: Developing Industry Applications
- Tableau Technical Series: Data Joins, LOD Calculations, Developing Dashboards & Machine Learning
- Tableau Server Series

# Data Science with Python



Learning Content from the perspective of a **Machine Learning  
Engineer @ Facebook**

## Salient features of the Course

**Course Workload** ~ 120 hours

**Videos duration** ~ 7 hours duration (on demand)

**Machine Learning Product Development With Python session:** 2 hours duration (online)

**Learner Level:** Beginners to Intermediate

**Interview Prep for Data Science with 200 + real interview Q&A:** Provided with A.I. Track

**More than 1000 IT & Business Professionals took this course since the year 2016**

**Total Companies** they represent > 50

**Total Countries** they represent > 10

Modules	Topics	Total Exercises + Case studies
11	Introduction to command line & GitHub	
12	Loading Data Into Python (Pandas Library Focus)	2
13	Cleaning and Saving Data Into Python (Pandas Library Focus)	2
14	Numerical Computation using Python ( Numpy Library Focus)	2
15	Inferential Statistics using Python	1
16	Data visualization using Python (Matplotlib Library Focus)	1
17	Summarizing Multivariate insights	1
18	A simple look at Machine Learning methods	1
19	EDA and Machine Learning Projects using Python	1



**Uber**



**McKinsey&Company**



## Data Science with Python

Become a Machine Learning Jedi

### Module 11: Introduction to command Line & Github (Optional Module with a guided workbook)

- Navigating the shell
- Paths and pwd
- Home Directory (~)
- Listing the Contents (ls)
- Changing Directories (cd)
- File Inspection (head and tail)
- Manipulating Files and Directories
- Creating Files (nano, emacs, vi, cat, >, and touch)
- Copying and Renaming Files (cp and mv)
- Making Directories (mkdir)
- Deleting Files and Directories (rm)
- Copying and Renaming Files (cp and mv)
- Flags and Wildcards
- Getting Help: Reading the Manual (man)
- Fundamanetals of Bash programming
- Developing your Project portfolio on **Github**
- How to work with Github

### Module 12: Loading data Into Python (Pandas Library Focus)

- Exploring Top 50 Jupyter Notebook Tricks and Tips
- Exploring Variables and Data Types in Python
- Data Structures in Python
- Flow control and Logic
- Writing loops in Python
- Most common List methods
- Most common Dictionary operations
- Most common String methods
- Dealing with Programming errors in Python for newbies
- Exploring Pandas Library functions

- Pandas Objects
- Manipulating Data frames (Indexing)
- Operations in Pandas
- Data Aggregations and Grouping
- Loading large and small data files into Python (.csv, .xlsx, SQL & JSON)
- Joining data using Python and SQL
- **Example datasets for practice:** Walmart and Flipkart

### **Module 13: Cleaning and Saving data in Python (Pandas Library Focus)**

- Examining the structure of the data frames
- Identifying key features
- Identifying missing and irrelevant data values
- Strategies to deal with missing data values in various domains like healthcare & finance
- Dealing with **None** Objects
- Dealing with **NaN** (not a number) values
- Exploring **isnull()**, **notnull()**, **dropna()**
- Techniques of data manipulation: Random Imputation, Mean, Median, Mode, Prediction based Imputation
- When to use which imputation method? Ordinal, Nominal, Ratio and Interval Data
- Handling Time data
- Applications of LDA algorithm in data imputation
- Using Series functions and loops to clean data
- Data Preparation: Binning, Standardization and Normalization techniques
- Exporting cleaned data into different file formats (.csv, .xlsx, SQL & JSON)
- **Example datasets for practice:** Uber and Enron

### **Module 14: Numerical Computation using Python ( Numpy Library Focus)**

- Computations on Numpy Arrays
- Boolean Logic in Numpy
- Fancy Indexing
- Sorting Arrays
- About Numpy Structured data
- Applications of Numpy in the Real world
- **Example datasets for practice:** Uber and Enron

## **Module 15: Data visualization using Python (Matplotlib & Seaborn Library Focus)**

- Examining the distribution of data using Histograms
- Digging depths of data distributions using Box Plots
- Visualizing categorical data using Bar plots
- Visualizing trends using Line charts
- Examining the relationships between numerical variables using Scatter Plots
- Examining the relations between multiple variables in a data frame using the correlation matrix

## **Module 16: Univariate, Bivariate & Multivariate Analysis using Python (Statsmodels Library and scikit learn Library focus)**

- Introduction to Probability concepts
- Descriptive & Inferential statistics
- Calculating Summary Statistics
- Applications of Probability Density Function
- Applications of Cumulative Distribution Function
- Interpretation of Median values
- Percentile and Quantiles
- Interpretation of Q-Q plots
- Identifying and handling outliers
- Calculating Correlations in multivariate data
- Interpreting Correlation Coefficient values
- Introduction to Principal Component Analysis
- Selecting Principal Components

## **Module 17: Hypothesis testing**

- How to Formulating Hypothesis?
- How do we test the Hypothesis?
- Null vs Alternative Hypothesis
- Critical values of alpha
- Applications of Student t-Test
- False Positives vs True Positives
- Applications of A/B Testing in Decision making

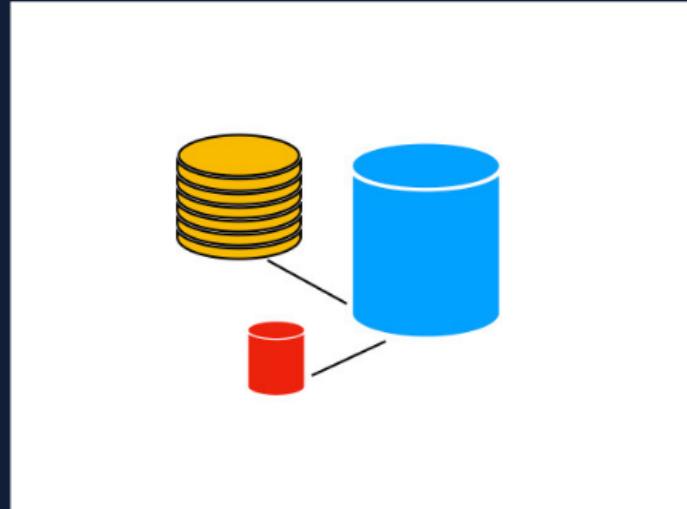
## **Module 18: Machine Learning: Developing AdHoc ML models in Healthcare Analytics domain (Statsmodels library and scikit learn Library focus)**

- Applications of Supervised and unsupervised learning in Healthcare Analytics
- Introduction to Healthcare Analytics domain
- Training & Testing Machine Learning Models
- Evaluating Machine Learning Model performance for Classification & Regression methods
- Selecting the right algorithms
- Developing Machine Learning Products in Healthcare Analytics
- An indepth introduction to applications of Scikit Learn
- Tweaking Machine Model Performances
- **Classification ML Project:** Identifying heart risk patients in a population over 10 years of duration
- **Classification ML Project:** Categorizing Future Risk through Healthcare insurance claims

## **Module 19: Kaggle Masters Class Machine Learning Focus**

- **AT&T:** Network anomaly detection using binomial and multinomial classifications
- **McKinsey:** Classify customers who are likely to purchase more Financial Products

# Data Science with SQL



Learning Content from the perspective of a **Data Engineer @ Accenture**

## Salient features of the Course

**Course Workload** ~ 50 hours

**Videos duration** ~ 5 hours duration (on demand)

**Learner Level:** Beginners to Advanced

**Interview Prep for Data Science with 200 + real interview Q&A:** Provided with A.I. Track

**More than 1000 IT & Business Professionals took this course since the year 2016**

**Total Companies** they represent > 50

**Total Countries** they represent > 10

Modules	Topics	Total Exercises + Case studies
20	Introduction to Databases	
	Data Extraction and query on XML & JSON	11
	Mastering SQL queries	11
	Data Cleaning Using SQL	11
	Advanced SQL queries	11
	Introduction to Hive	1
	Introduction to NoSQL	1
	Develop Databases using Web Scrapped data	1
	SQL Interview Prep	1

# Data Science with SQL

Become a SQL Jedi

## Module 20: Databases for Data science

### Part 1

- Introduction to Relational Databases
- Basics of Querying Relational Databases
- Introduction to XML Data
- Querying XML
- Introduction to JSON Data
- Querying JSON Data

### Part 2

- Introduction to SQL Query Language
- How to use the SELECT Statement?
- Subqueries using WHERE statement
- Subqueries using FROM and SELECT
- Data aggregation in SQL
- Dealing with NULL Values
- Data Cleaning procedures in SQL
- Data Modification Statements

### Part 3

- Restricting and sorting data
- Single row functions
- Multiple row functions
- SQL Joins
- Sub-Queries
- Creating & Managing tables
- Set operators
- Pseudo columns
- SQL Constraints
- Sequences
- Views
- Indexes
- Synonyms
- Oracle date time functions
- Advanced sub queries

- Advanced PL/SQL
- More on Loops; Cursors; Stored Procs; Function; Triggers

#### Part 4

- Working with Hive ecosystem
- NoSQL Systems
- NoSQL Motivation
- NoSQL Overview

#### Part 5

**Project:** Develop a **Data Scientist's First Database** using Web Scrapped competitive intelligence data from **Flipkart & Amazon**. Extract data about top 30 Smartphones with focus on Product Pricing and Product positioning.

# Develop Data Pipelines for Data Science



Learning Content from the perspective of a **Data Engineer @ Accenture**

## Salient features of the Course

**Course Workload** ~ 50 hours

**Videos duration** ~ 5 hours duration (on demand)

**Learner Level:** Beginners to Advanced

**Interview Prep for Data Science with 200 + real interview Q&A:** Provided with A.I. Track

**More than 1000 IT & Business Professionals took this course since the year 2016**

**Total Companies** they represent > 50

**Total Countries** they represent > 10

Modules	Topics	Total Exercises + Case studies
21	Exploring UI of Alteryx Designer	1
22	Advanced Data Analysis and Product Development in Data Science	1



## Data Science with Alteryx

Become a Data Product Design Expert

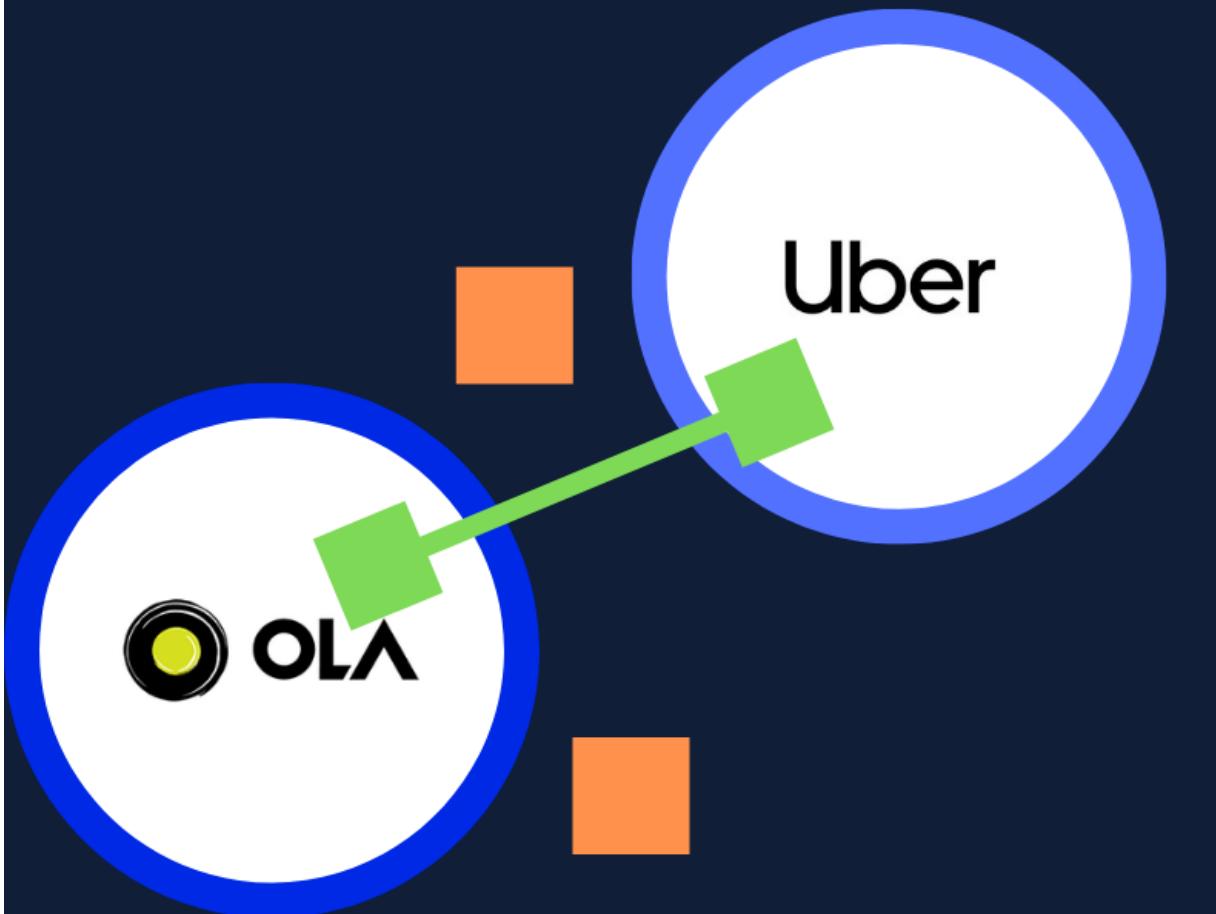
### Module 21: Exploring UI of Alteryx Designer

- Understanding data types
- Inputting structured and **unstructured data** into the workflow
- Manual Data entry techniques
- Blending Data and Joins
- Viewing data
- Filtering data
- Selecting and Sorting data
- Exploring Vlookups
- Separating data in columns and rows
- Data Cleansing
- Sampling Data
- Writing Expressions
- Writing conditional statements
- Applying String functions
- Applying numeric functions
- Evaluating null and empty values
- Calculations with dates and Times

### Module 22: Advanced Data Analysis and Product Development in Data Science

- Using Spatial functions
- Viewing Spatial Objects
- Understanding Spatial Objects
- Creating Spatial Objects
- Constructing Trade areas
- Splitting Spatial Objects
- Calculating Distances
- Relating Spatial objects
- Identifying nearest objects
- Spatially Joining Data
- Using Analytics Applications
- Designing App Interfaces
- Specifying data input
- Chaining down Analytics product

# Developing Data Products an Interview Perspective



## Building Data Science Products

Ace Data Science Problem solving Interviews

**Module 23:** Exploring System design and Architecture of Uber Machine Learning System

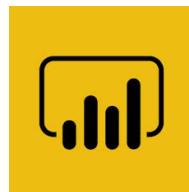
**Module 24:** Exploring System design and Architecture of OLA Machine Learning System

**Module 25:** Exploring System design and Architecture of Amazon Machine Learning System

**Module 26:** Exploring System design and Architecture of Airbnb Machine Learning System

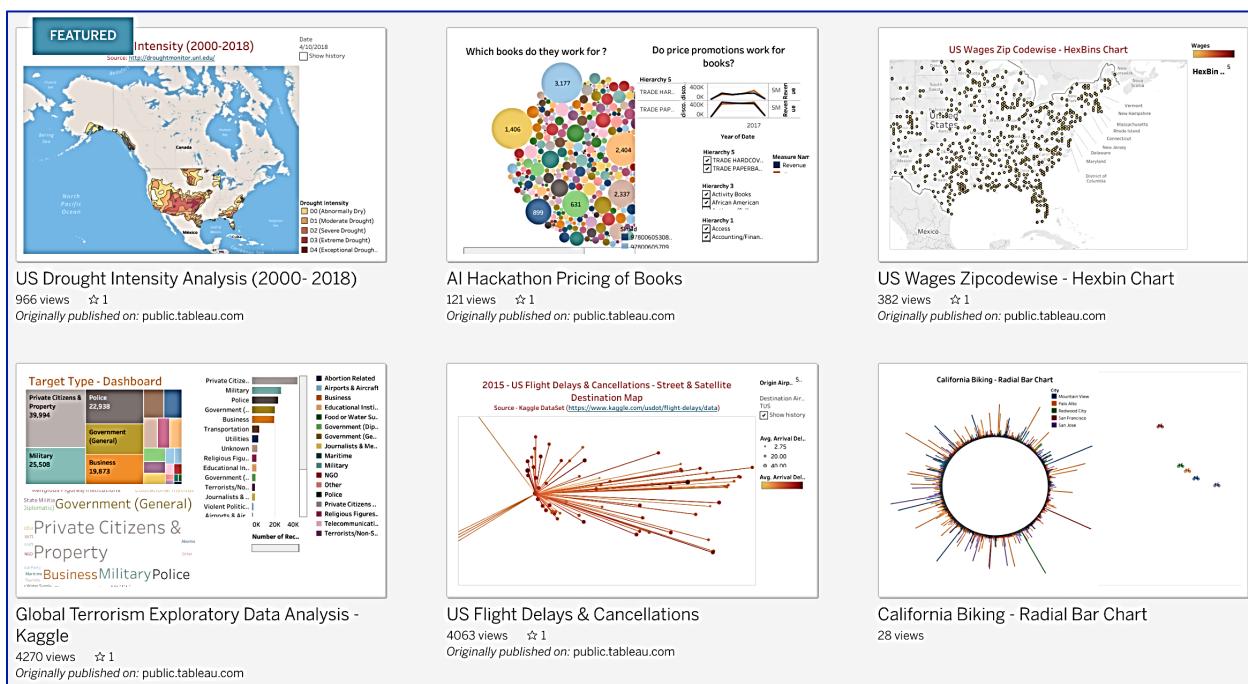
**Module 26:** Exploring System design and Architecture of Whatsapp Machine Learning System

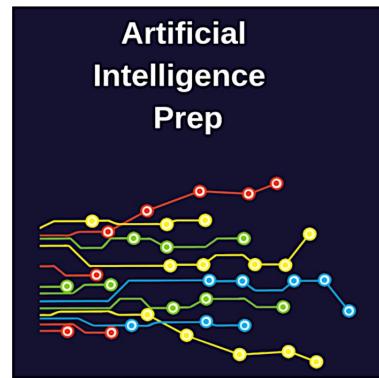
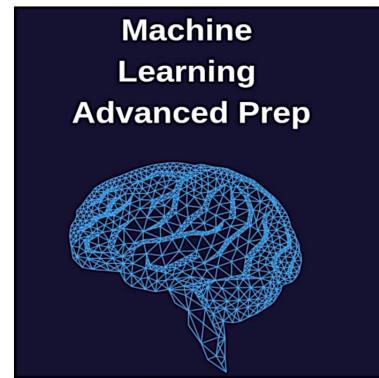
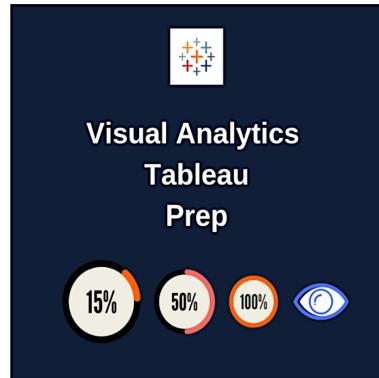
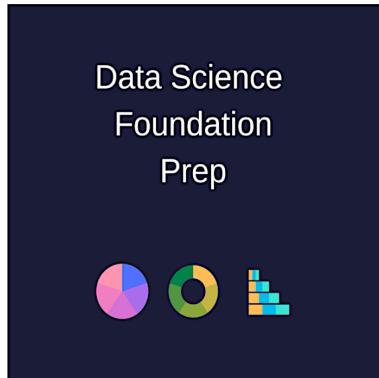
# Free Course Addons in 2019



# Focus on your learning outcomes

- Machine Learning Projects Portfolio
  - Crack Interviews
  - Have a successful career transition





**500 +** Real World  
Technical Round Interview  
Questions

**30 +** Real interview  
Technical Tasks

# India's Most successful Career Transition Program into Data Science & Machine Learning since 2016



**CDP** 

Career Development Program  
2019



Resume Development  
Workshop 2019



Career Mapping for Future  
Data Scientists 2019

info@datascienceprep.co

