DATA SCIENCE

CURRICULUM

8096096564

https://www.linkedin.com/in/rajendra-prasad-vadla/

rp.statistician@gmail.com

Introduction to Data Science Overview

- What is Data Science?
- Why Data Science?
- Where it is used?
- Career Opportunities
- About TEKS Academy
- Learning Journey

Installation of Anaconda
Overview on Jupiter notebook

Module 01: Basic Python

Python Programming

Introduction to Python

- Overview of Python and its features
- Installing Python
- Python IDEs
- Writing and executing Python programs
- Understanding Python's interactive mode and script mode

Python Basics

- Python syntax and indentation
- Python variables and data types (int, float, str, bool)
- Input/output functions (input(), print())
- Comments in Python

Data Structures in Python

- Lists:
  - Creating and manipulating lists
  - List functions (append(), insert(), remove(), slicing)
- Tuples:
  - Difference between lists and tuples
  - Accessing elements in a tuple

- Sets:
  - Creating sets, set operations
  - union(), intersection(), difference()
- Dictionaries:
  - Key-value pairs
  - Accessing and updating dictionaries
  - Common dictionary methods (get(), items(), keys(), values())

Control Structures

- Conditional statements (if, elif, else)
- Loops:
  - for loop
  - while loop
  - break, continue, and pass statements
- Logical and comparison operators

Functions

- Defining functions in Python
- Function arguments and return values
- Default and keyword arguments
- *args and **kwargs
- Lambda functions

Exception Handling

- Errors in Python (syntax and runtime errors)
- try, except, finally blocks
- Raising exceptions using raise

Modules and Packages

- Importing modules (import, from...import)
- Standard Python libraries (math, random, os, sys, datetime)
- Creating and using your own modules

8096096564 www.linkedin.com/in/rajendra-prasad-vadla rp.statistician@gmail.com

File Handling

- Reading from and writing to files (open(), read(), write())
- Working with file modes (r, w, a, rb, wb)
- Handling file exceptions

Object-Oriented Programming (OOPs) Basics

- Introduction to classes and objects
- Defining a class and object

Introduction to Python Libraries

- Using Python libraries such as NumPy, Matplot and Pandas
- Simple data analysis examples using these libraries

Module 02: Statistics and Probability

Introduction to Statistics

- Definition and importance of statistics in data science
- Types of data:
  - Numerical (Discrete and Continuous)
  - Categorical (Nominal and Ordinal)
- Types of statistics:
  - Descriptive Statistics
  - Inferential Statistics
- Population vs. Sample
- Data Collection Techniques (Surveys, Experiments, Observations)

Descriptive Statistics

- Measures of Central Tendency:
  - Mean, Median, Mode
- Measures of Dispersion:
  - Range, Variance, Standard Deviation
  - Interquartile Range (IQR)
- Shape of Data Distribution:
  - Skewness and Kurtosis

- Data Visualization Techniques:
  - Bar charts, Histograms, Pie charts
  - Box plots, Scatter plots
  - Heatmaps

Probability Basics

- Definition of Probability
- Types of Probability:
  - Classical, Empirical, and Subjective Probability
- Probability Rules:
  - Addition and Multiplication rules
  - Conditional Probability
- Independent and Dependent Events
- Bayes' Theorem and Applications

Random Variables and Probability Distributions

- Definition of Random Variables (Discrete and Continuous)
- Probability Distribution of Random Variables
- Expectation, Variance, and Standard Deviation of Random Variables
- Cumulative Distribution Function (CDF)
- Probability Density Function (PDF)

Sampling and Sampling Distributions

- Importance of Sampling in Data Science
- Types of Sampling Methods:
  - Simple Random Sampling
  - Stratified Sampling
  - Cluster Sampling
  - Systematic Sampling
- Sampling Distribution of the Sample Mean
- Central Limit Theorem (CLT)
- Standard Error

8096096564 www.linkedin.com/in/rajendra-prasad-vadla rp.statistician@gmail.com

Inferential Statistics

- Point Estimation vs. Interval Estimation
- Confidence Intervals for Means and Proportions
- Hypothesis Testing:
  - Null and Alternative Hypotheses
  - Type I and Type II Errors
  - P-value and Significance Level
- Z-test, T-test, and Chi-square test
- ANOVA (Analysis of Variance)
- Power of a Test

Correlation and Regression

- Covariance and Correlation
- Pearson Correlation Coefficient

Module 03: Data Cleaning with Pandas and Numpy

Introduction to Data Cleaning

- Importance of Data Cleaning in Data Science
- Common Data Quality Issues:
  - Missing Data
  - Duplicate Data
  - Incorrect Data Types
  - Inconsistent Data
- Overview of Pandas and NumPy for Data Cleaning

Introduction to Pandas and NumPy

- Installing and Setting Up Pandas and NumPy
- Overview of Pandas DataFrames and Series
- Overview of NumPy Arrays and Basic Operations
- Importing Data using Pandas:
  - CSV, Excel, and JSON files
- Data Inspection:
  - head(), info(), describe(), shape, and dtypes

Handling Missing Data

- Identifying Missing Data:
  - isnull(), notnull(), isna(), sum()
- Filling Missing Values:
  - Using fillna() and ffill(), bfill()
  - Filling with Mean, Median, Mode
  - Interpolation Techniques
- Dropping Missing Values:
  - dropna() function and its parameters
- Replacing Values using replace()

Handling Duplicate Data

- Identifying Duplicate Rows:
  - duplicated(), drop_duplicates()
- Removing Duplicate Rows and Columns
- Dealing with Duplicate Values based on Conditions

Data Type Conversion

- Checking Data Types with dtypes
- Converting Data Types using astype():
  - Converting between integers, floats, and strings
- Handling Date and Time Data with Pandas:
  - Converting to datetime using to_datetime()
  - Extracting date and time components (day, month, year, hour, minute)

Handling Inconsistent Data

- String Manipulation:
  - Cleaning text data using str methods
  - Case conversion (lower(), upper())
  - Removing whitespace and special characters
  - Replacing substrings in text data
- Dealing with Inconsistent Labels:
  - Renaming columns with rename()
  - Standardizing labels

8096096564 www.linkedin.com/in/rajendra-prasad-vadla rp.statistician@gmail.com

Handling Outliers

- Identifying Outliers:
  - Using statistical techniques (IQR, Z-score)
  - Visualizing outliers with Boxplots and Histograms
- Treating Outliers:
  - Capping, Flooring, and Winsorization
  - Removing or transforming outliers

Data Transformation with Pandas

- Applying Functions to Data using apply(), map(), and applymap()
- Lambda Functions for Custom Operations
- Creating New Columns from Existing Data
- Grouping and Aggregating Data:
  - groupby(), aggregate(), transform()
- Pivoting and Unpivoting Data with pivot(), melt()

Working with Large Datasets

- Handling Large Data with NumPy:
  - Efficient data storage with NumPy arrays
- Loading and Manipulating Large Files with Pandas:
  - Chunking large datasets
  - Memory optimization techniques (downcasting)
- Using Dask for large-scale DataFrames

Merging and Joining DataFrames

- Concatenating DataFrames with concat()
- Merging DataFrames with merge():
  - Types of Joins (Inner, Outer, Left, Right)
- Combining DataFrames using join()

Data Cleaning with NumPy

- Introduction to NumPy Arrays for Data Cleaning
- Element-wise Operations on Arrays

8096096564 www.linkedin.com/in/rajendra-prasad-vadla rp.statistician@gmail.com

- Handling Missing Values in NumPy:
  - np.nan, np.isnan(), and np.nan_to_num()
- Using np.where() for Conditional Data Cleaning
- Efficient Data Filtering using Boolean Indexing

Module 04: Normalization preprocessing using Scikit learn

Introduction to Data Normalization

- Importance of Data Preprocessing in Data Science
- Why Normalization is Needed:
  - Impact on Machine Learning Algorithms
  - Impact on Distance-Based Algorithms (KNN, K-means, etc.)
- Difference between Normalization and Standardization
- Overview of Scikit-learn Library

Overview of importing datasets using Scikit-learn and Seaborn

- Introduction to Sklearn's datasets module
- Loading Data using Scikit-learn and Seaborn:
  - Datasets (Iris, Boston Housing, Penguins, titanic, etc..,)
  - Importing and converting data to Scikit-learn's format
- Data Inspection Techniques

Types of Normalization Techniques

- Min-Max Scaling (Normalization):
  - Definition and Formula
  - Scaling values between 0 and 1
  - When to use Min-Max Scaling
- Z-Score Standardization (Standard Scaling):
  - Definition and Formula
  - Mean and Standard Deviation-based scaling
  - When to use Standardization
- MaxAbsScaler:
  - Scaling data by dividing by the maximum absolute value
  - Suitable for data that is already centered at 0

- RobustScaler:
  - Scaling using the median and IQR to handle outliers
  - When to use RobustScaler

Data Transformation with Scikit-learn

- Importing and Using MinMaxScaler
- Importing and Using StandardScaler
- Importing and Using MaxAbsScaler
- Importing and Using RobustScaler
- Applying Multiple Transformations using Pipeline

Applying Normalization to Different Data Types

- Handling Numerical Features
- Normalizing Categorical Data (One-Hot Encoding, Label Encoding)
- Dealing with Mixed Data Types:
  - Custom Pipelines for Handling Heterogeneous Data

Normalizing and Standardizing Data with Scikit-learn

- Using normalize() Function:
  - L1 and L2 Normalization Techniques
- Data Preparation for Machine Learning Algorithms
- Combining Normalization with Feature Engineering

Handling Outliers in Normalization

- Effects of Outliers on Normalization Techniques
- Scaling Data with Outliers using RobustScaler
- Data Transformation using Log, Square Root, and Box-Cox Transformations

Feature Scaling for Machine Learning Models

- Feature Scaling in Regression Algorithms
- Normalization in Distance-Based Algorithms:
  - K-Nearest Neighbors (KNN)
  - K-Means Clustering

8096096564 www.linkedin.com/in/rajendra-prasad-vadla rp.statistician@gmail.com

- Importance of Scaling in Gradient-Based Algorithms:
  - Logistic Regression
  - Support Vector Machines (SVM)

Using Pipelines for Data Preprocessing

- Overview of Scikit-learn Pipelines
- Creating Custom Pipelines for Normalization and Preprocessing
- Integrating Scaling and Normalization in Machine Learning Pipelines
- Cross-validation with Preprocessing Pipelines

Dealing with Imbalanced Data

- Handling Class Imbalances Before Normalization
- Scaling Imbalanced Data:
  - Over-sampling and Under-sampling Techniques
- Applying Normalization Post-Sampling

Data Normalization for Deep Learning

- Normalization Techniques for Neural Networks
- Batch Normalization
- Importance of Scaling Inputs in Deep Learning

Module 05: Explorative Data Analysis with Visualization

Introduction to Exploratory Data Analysis (EDA)

- Importance of EDA in Data Science
- Goals of EDA: Detecting patterns, identifying anomalies, and hypothesis formulation
- Overview of Tools for EDA:
  - Pandas for data manipulation
  - Matplotlib, Seaborn, and Plotly for data visualization

Introduction to Data Visualization

- Overview of Visualization Tools: Matplotlib, Seaborn, and Plotly
- Basic Plot Types:
  - Line Plot, Bar Plot, Scatter Plot

Univariate Analysis with Matplotlib & Seaborn

- Visualization of Single Variables
- Numerical Data:
  - Histograms, Boxplots, Violin Plots
- Categorical Data:
  - Bar Charts, Count Plots
- Kernel Density Estimation (KDE) Plots

Bivariate Analysis with Matplotlib & Seaborn

- Exploring Relationships between Two Variables
- Scatter Plots for Numerical Data
- Boxplots and Violin Plots for Categorical vs. Numerical Data

Multivariate Analysis with Matplotlib & Seaborn

- Visualizing Multiple Variables
- Pair Plots, Joint Plots, and Facet Grids
- 3D Scatter Plots and Bubble Charts
- Correlation Matrices and Heatmaps

Customizing with Matplotlib

- Customizing Plots: Titles, Labels, Legends, Grids
- Subplots and Figure Layouts
- Saving and Exporting Figures

Time Series Visualization

- Line Plots for Time Series Data
- Rolling Statistics and Moving Averages
- Time Series Decomposition for Trend and Seasonality

 8096096564  www.linkedin.com/in/rajendra-prasad-vadla  rp.statistician@gmail.com

Module 06: Machine Learning -Supervised Learning

- Overview of Machine Learning

Introduction to Supervised Learning

- Introduction to Supervised Learning
- Difference between Supervised and Unsupervised Learning
- Common Applications of Supervised Learning
- Overview of Training and Testing Process:
- Labels and Features, Train-Test Split

Data Preprocessing for Supervised Learning

- Data Cleaning and Feature Scaling
- Handling Missing Values
- Encoding Categorical Data
- Feature Engineering and Selection
- Data Normalization and Standardization

Linear Regression

- Introduction to Regression Problems
- Simple Linear Regression
- Multiple Linear Regression
- Assumptions of Linear Regression
- Evaluating Regression Models:
- Mean Squared Error (MSE), Root Mean Squared Error (RMSE), R-Squared Value

Logistic Regression

- Introduction to Classification Problems
- Logistic Regression for Binary Classification
- Sigmoid Function and Decision Boundaries
- Evaluating Classification Models:
- Confusion Matrix, Precision, Recall, F1-Score, ROC Curve, AUC

Decision Trees

- Splitting Criteria: Gini Index, Entropy
- Splitting Criteria: Gini Index, Entropy
- Overfitting and Pruning Techniques
- Pros and Cons of Decision Trees

Bagging and Random Forest

- Introduction to Ensemble Learning
- Working of Random Forest Algorithm
- Feature Importance in Random Forest
- Hyperparameter Tuning: Number of Trees, Max Depth

Support Vector Machines (SVM)

- Introduction to Support Vector Machines
- Linear and Non-linear SVM
- Kernel Trick: Polynomial, RBF Kernel
- Margin Maximization and Support Vectors
- Hands-on: Handwritten Digit Classification

k-Nearest Neighbors (k-NN)

- Introduction to k-NN Algorithm
- Choosing the Optimal Value of k
- Distance Metrics: Euclidean, Manhattan
- Pros and Cons of k-NN
- Hands-on: Classifying Iris Dataset

Naive Bayes Classifier

- Introduction to Naive Bayes Algorithm
- Types of Naive Bayes: Gaussian, Multinomial, Bernoulli
- Assumptions and Applications of Naive Bayes
- Hands-on: Spam Detection using Naive Bayes

8096096564 www.linkedin.com/in/rajendra-prasad-vadla rp.statistician@gmail.com

Module 07: Ensemble Models

Introduction to Ensemble Learning

- Overview of Ensemble Learning
- Benefits of Ensemble Methods:
  - Reducing Overfitting
  - Improving Prediction Accuracy
- Types of Ensemble Learning:
  - Bagging, Boosting, Stacking

Bagging Techniques

- Introduction to Bagging:
  - Definition and Concept
  - How Bagging Reduces Variance
- Decision Tree Ensembles:
  - Random Forest Algorithm
  - Building and Tuning Random Forest Models
- Out-of-Bag Error Estimation

Boosting Techniques

- Introduction to Boosting:
  - Definition and Concept
  - How Boosting Reduces Bias
- Common Boosting Algorithms:
  - AdaBoost
  - Gradient Boosting Machines (GBM)
  - XGBoost: Features and Implementation
  - LightGBM and CatBoost: Differences and Use Cases

Stacking and Blending

- Introduction to Stacking:
  - Definition and Concept
  - Combining Multiple Models

8096096564 www.linkedin.com/in/rajendra-prasad-vadla rp.statistician@gmail.com

- Building a Stacked Model:
  - Choosing Base Learners
  - Using a Meta-learner
- Practical Implementation of Stacking

Module 08: Machine Learning – Unsupervised

Introduction to Unsupervised Learning

- Overview of Unsupervised Learning
- Difference Between Supervised and Unsupervised Learning
- Applications of Unsupervised Learning
- Overview of Clustering and Dimensionality Reduction

k-Means Clustering

- Introduction to Clustering
- Working of k-Means Algorithm
- Evaluating Clusters:
  - Elbow Method to Find Optimal Number of Clusters
  - Silhouette Score

Hierarchical Clustering

- Introduction to Hierarchical Clustering
- Agglomerative vs. Divisive Clustering
- Dendrograms and Linkage Methods
- Pros and Cons of Hierarchical Clustering

DBSCAN (Density-Based Spatial Clustering)

- Introduction to DBSCAN Algorithm
- Concept of Density Reachability and Density Connectivity
- Identifying Outliers and Noise in DBSCAN
- Choosing Parameters: Epsilon, MinPts

8096096564 www.linkedin.com/in/rajendra-prasad-vadla rp.statistician@gmail.com

Principal Component Analysis (PCA)

- Introduction to Dimensionality Reduction
- Working of PCA Algorithm
- Eigenvalues and Eigenvectors in PCA
- Explained Variance and Choosing Number of Components
- Visualizing PCA for High-Dimensional Data

t-SNE (t-Distributed Stochastic Neighbor Embedding)

- Introduction to t-SNE for Non-linear Dimensionality Reduction
- Visualizing High-Dimensional Data in 2D/3D
- Differences between PCA and t-SNE

Association Rule Learning (Apriori Algorithm)

- Introduction to Market Basket Analysis
- Support, Confidence, and Lift
- Apriori Algorithm for Finding Association Rules
- Applications of Association Rules in Retail and E-commerce

Anomaly Detection

- Introduction to Anomaly Detection
  - Techniques for Identifying Outliers
- Isolation Forest, One-Class SVM
- Applications of Anomaly Detection in Fraud Detection and Network Security

Module 09: Time Series Analysis

Introduction to Time Series Analysis

- Characteristics of Time Series Data
- Components of Time Series:
  - Trend, Seasonality, Noise
- Importance of Time Series Analysis in Data Science

8096096564 www.linkedin.com/in/rajendra-prasad-vadla rp.statistician@gmail.com

Time Series Decomposition

- Decomposing Time Series into Components
- Seasonal and Trend Decomposition using STL (Seasonal-Trend decomposition using Loess)
- Visualizing Decomposed Components

Time Series Forecasting Methods

- Introduction to AR & MA Models :
  - AR Model
  - MA Model
  - ARMA Model
  - ARIMA Model

Evaluating Time Series Models

- Metrics for Time Series Forecasting:
  - MAE, MSE, RMSE, MAPE
- Train-Test Split for Time Series Data
- Walk-Forward Validation

Module 10: Deep Learning

Introduction to Deep Learning

- Overview of Artificial Intelligence, Machine Learning, and Deep Learning
- Historical Context and Evolution of Deep Learning
- Key Differences between Traditional ML and Deep Learning
- Applications of Deep Learning in Data Science

Foundations of Neural Networks

- Introduction to Neurons and Perceptrons
- Structure of a Neural Network
- Activation Functions: Sigmoid, ReLU, eLU, Leaky ReLU, Tanh, Softmax, Softplus
- Forward Propagation and Loss Calculation
- Understanding Cost Functions

8096096564 www.linkedin.com/in/rajendra-prasad-vadla rp.statistician@gmail.com

Training Neural Networks

- Backpropagation Algorithm
- Gradient Descent and its Variants:
- Stochastic Gradient Descent (SGD)
- Adam, RMSprop
- Learning Rate and its Importance
- Overfitting and Underfitting
- Techniques to Prevent Overfitting: Regularization, Dropout

Deep Neural Networks (DNNs)

- Constructing Deep Neural Networks
- Importance of Depth and Width in Networks
- Vanishing and Exploding Gradients Problem
- Batch Normalization

Practical Implementation with TensorFlow and Keras

- Setting Up the Environment for TensorFlow
- Building and Training Models using Keras
- Handling Datasets with TensorFlow Datasets (TFDS)
- Model Evaluation and Metrics: Accuracy, Precision, Recall, F1-score
- Saving and Loading Models

Practical Implementation with PyTorch

- Setting Up the Environment for PyTorch
- Building Neural Networks with PyTorch
- Data Handling and Augmentation in PyTorch
- Model Training and Evaluation in PyTorch
- Using Pretrained Models in PyTorch

8096096564 www.linkedin.com/in/rajendra-prasad-vadla rp.statistician@gmail.com

Module 11: Natural Language Processing (NLP) with NLTK

- Introduction
- Text Preprocessing Techniques
- Stemming, Limitation, Stop words,
- TFIDE, Word2 vec
- Word Embedding
- Word Cloud.

Deployment using Streamlit

Module 12: SQL Data Base

Introduction to SQL

- History and evolution of SQL
- SQL vs NoSQL
- Types of databases (RDBMS, column-based, key-value, etc.)
- Database concepts: Tables, Rows, Columns, Relationships

SQL Data Types

- Numeric types (INT, FLOAT, DECIMAL)
- Character types (CHAR, VARCHAR, TEXT)
- Date and time types (DATE, TIME, TIMESTAMP)
- Boolean types
- BLOB (Binary Large Object)

Database Design

- Normalization (1NF, 2NF, 3NF, BCNF)
- Denormalization
- Primary keys, foreign keys, and unique keys
- Indexing
- Constraints (NOT NULL, DEFAULT, UNIQUE, CHECK)

Basic SQL Queries

- SELECT statement
- WHERE clause and logical operators (AND, OR, NOT)
- ORDER BY clause
- LIMIT and OFFSET clauses
- DISTINCT keyword

SQL Functions

- Aggregate functions (COUNT, SUM, AVG, MIN, MAX)
- Scalar functions (UPPER, LOWER, LENGTH, ROUND)
- Date functions (NOW, CURDATE, DATE_ADD, DATE_SUB)

Joins in SQL

- INNER JOIN
- LEFT JOIN (or LEFT OUTER JOIN)
- RIGHT JOIN (or RIGHT OUTER JOIN)
- FULL OUTER JOIN
- CROSS JOIN
- Self joins

Subqueries and Nested Queries

- Single-row subqueries
- Multi-row subqueries
- Correlated subqueries
- EXISTS and NOT EXISTS clauses

Set Operations

- UNION and UNION ALL
- INTERSECT
- EXCEPT (or MINUS)

8096096564 www.linkedin.com/in/rajendra-prasad-vadla rp.statistician@gmail.com

Data Manipulation Language (DML)

- INSERT statement
- UPDATE statement
- ADD statement
- DELETE statement
- TRUNCATE statement

Data Definition Language (DDL)

- CREATE TABLE
- ALTER TABLE (add, modify, drop columns)
- DROP TABLE
- CREATE VIEW, DROP VIEW

Constraints in SQL

- PRIMARY KEY constraint
- FOREIGN KEY constraint
- UNIQUE constraint
- CHECK constraint
- DEFAULT constraint

Transactions in SQL

- ACID properties (Atomicity, Consistency, Isolation, Durability)
- COMMIT and ROLLBACK
- SAVEPOINT
- Transaction isolation levels (READ UNCOMMITTED, READ COMMITTED, REPEATABLE READ, SERIALIZABLE)

Indexes in SQL

- Purpose of indexes
- Types of indexes (single-column, multi-column)
- Unique and non-unique indexes
- Full-text index
- Index performance considerations

8096096564 www.linkedin.com/in/rajendra-prasad-vadla rp.statistician@gmail.com

SQL Views

- Creating views
- Updating views
- Dropping views
- Advantages and limitations of views
- SQL Trigger

Stored Procedures and Functions

- Creating stored procedures
- IN, OUT, and INOUT parameters
- Creating user-defined functions
- Differences between stored procedures and functions

Module 13: Power BI

Introduction to Power BI

- Overview of Business Intelligence and Data Visualization
- Importance of Power BI in Data Science
- Components of Power BI:
  - Power BI Desktop, Power BI Service, Power BI Mobile
- Installation and Setup of Power BI Desktop

Getting Started with Power BI Desktop

- Interface Overview: Ribbon, Fields Pane, Visualizations Pane
- Importing Data:
  - Connecting to various data sources (Excel, SQL, Web, etc.)
- Understanding Data Types and Basic Data Profiling

Data Transformation with Power Query

- Introduction to Power Query Editor
- Data Cleaning Techniques:
  - Removing duplicates, filtering rows, and changing data types
- Merging and Appending Queries
- Creating Custom Columns and Calculated Fields
- Handling Missing Values

Data Modeling in Power BI

- Understanding Relationships: One-to-One, One-to-Many, Many-to-Many
- Creating and Managing Relationships between Tables
- Using Star Schema and Snowflake Schema for Data Models
- Introduction to Data Hierarchies

DAX (Data Analysis Expressions) Basics

- Introduction to DAX: What It Is and Why It's Important
- Creating Calculated Columns and Measures
- Basic DAX Functions:
  - SUM, AVERAGE, COUNT, DISTINCTCOUN
- Time Intelligence Functions:
  - YTD, QTD, MTD calculations

Data Visualization Techniques

- Creating Basic Visualizations:
  - Stacked Column charts, Line charts, Pie charts,Donut Chart, Ribbon Plot, Tables, and Matrix
- Advanced Visualizations:
  - Treemaps, Waterfall charts, Scatter plots, Maps
- Custom Visualizations from Power BI Marketplace
- Best Practices for Data Visualization Design

Creating Interactive Reports and Dashboards

- Designing Interactive Reports:
  - Using slicers, filters, and drill-through functionality
- Creating Bookmarks and Buttons for Navigation
- Tips for Effective Dashboard Design
- Publishing Reports to Power BI Service

8096096564 www.linkedin.com/in/rajendra-prasad-vadla rp.statistician@gmail.com

Module 14: Data Summarization, Charts & Formatting

- Getting data from CSV files, databases, workbooks, webpages
- Max, Min with IF and IFS, CountIFS
- Sum, Product, Sumproduct, Average, Standard Deviation, Variance
- LOOKUP,VLOOKUP, HLOOKUP
- Various types of Charts – pie chart, Column chart, line chart, Scatter Plot
- Changing Font, Data Type of column, Conditional Formatting, Format painter
- Alignment Techniques – Merging and Wrapping
- Data Summarization: Pivot Table
- Cleaning data, extracting data from multiple sources
- Filters, Sorting, Concatenation, Merging

8096096564

www.linkedin.com/in/rajendra-prasad-vadla

rp.statistician@gmail.com