

VEER NARMAD SOUTH GUJARAT UNIVERSITY, SURAT

SYLLABUS

Title of Course	AI-Driven Statistical Analytics		
Credit	Theory: 2	Practical: 2	Total: 4
Course Content	Unit – 1: Foundations of Quantitative Analysis & Computational Environments Theoretical Component • Introduction to Modern Analytics Data Pipelines: The transition from desktop statistics to enterprise-scale automated computational analytics frameworks. • Core Statistical Principles Revisited: Types of variables (nominal, ordinal, interval, ratio); understanding population parameters vs. sample statistics; the Law of Large Numbers and Central Limit Theorem. • Descriptive Measures and Distributions: Mathematical formulations of central tendency (mean, median, mode), robust dispersion metrics (variance, standard deviation, interquartile range, skewness, kurtosis); probability density functions for Gaussian, Binomial, and Poisson variations. • The Computational Environment Setup: Architecture of integrated development environments; standard script execution models; understanding dependency tree structures and workspace management. Practical & Laboratory Component • Workspace Engineering: Initializing terminal environments; launching and navigating Jupyter Notebook systems and VS Code local directories. • Python Syntax Core Implementations: Declaring structural data containers (Lists, Dictionaries, Tuples, Sets); building iterative validation pipelines using loops (for, while) and functional operations (if-else). • Functional Programming Design: Writing modular, cleanly documented user-defined functions with parameterized argument mapping and return pipelines. • Mathematical Grid Operations: Utilizing the NumPy library to initialize matrices, run multidimensional vector calculations, and perform linear algebra translations. Unit – 2: Advanced Exploratory Data Analysis & Inference Engines Theoretical Component • Data Ingestion & Structural Management Theory: Relational structures vs. unstructured text frames; working with sparse matrices, missing row values, and unaligned column datatypes. • Exploratory Data Analysis (EDA) Paradigms: Principles of tabular structure analysis; tracking underlying distributions; identifying skewness, outliers, and data collection noise. • Inferential Testing Frameworks: Formulation of Null (H_0) and Alternative (H_1) hypotheses; understanding Type-I (α)		

and Type-II (β) errors; the mechanics of the p-value and statistical confidence limits.

- Parametric vs. Non-Parametric Frameworks: Theoretical constraints, mathematical assumptions, and mathematical boundaries governing Student's t-tests, Analysis of Variance (ANOVA), and Chi-Square tests of independence.

Practical & Laboratory Component

- Data Transformation Workflows (Pandas): Parsing external assets (.csv, .xlsx, .json); structural querying via boolean vector filters; row grouping, matrix indexing, and running aggregate mutations.
- Data Cleaning Implementations: Automating systematic imputation strategies for structural blanks; using statistical z-scores and interquartile ranges (IQR) to detect and manage outliers.
- Visual Storyboarding Scripts: Writing customized scripts using Matplotlib and Seaborn to display feature correlations (heatmaps, scatter plot grids, pairplots, boxplots).
- Computational Inference Modeling: Executing hypothesis test scripts via the SciPy and Statsmodels ecosystems; processing ANOVA tables and parsing calculated test statistics against alpha parameters.

Unit – 3: Supervised Machine Learning & Statistical Optimization

Theoretical Component

- Algorithmic Learning Principles: Supervised machine learning paradigms; mathematical formulation of optimization vectors; loss function minimizing protocols.
- Linear and Curvilinear Regression Models: Ordinary Least Squares (OLS) optimization matrices; intercept and weight derivations; multi-collinearity checks (Variance Inflation Factor); residual validation testing (homoscedasticity, autocorrelation).
- Probabilistic Classification Models: The Sigmoid function activation mechanics; odds ratio calculations; log-likelihood maximization functions in binary classification pipelines.
- Non-Parametric Structural Learning Models: Recursive binary partitioning theory; Gini impurity minimization algorithms and entropy reductions used to build Decision Tree topologies.

Practical & Laboratory Component

- Pipeline Engineering via Scikit-Learn: Automated transformation steps for column data; converting categorical records into vector fields via One-Hot Encoding and ordinal map transformations.
- Regression Model Training: Setting up multi-variable linear regressions; extracting optimization parameters; evaluating prediction success through Mean Squared Error (MSE) and Adjusted R^2 scores.
- Classification Engine Implementation: Deploying Logistic Regression classification models and tracking model boundaries; adjusting binary threshold parameters dynamically.
- Tree Architecture Evaluations: Building Decision Tree regression and classification engines; configuring tree parameters (max depth, min samples split) to control model growth.

Unit – 4: Unsupervised Paradigms & Pattern Mining Systems

Theoretical Component

- Unsupervised Framework Principles: Pattern discovery without label parameters; clustering mechanics vs. dimensionality reduction math.
- Centroid-Based Partitioning Algorithms: Distance-metric mechanics (Euclidean, Manhattan, Minkowski metrics); iterative convergence math behind the K-Means clustering algorithm.
- Cluster Validation Frameworks: Within-Cluster Sum of Squares (WCSS) optimization curves; calculating individual silhouette coefficient values to measure separation quality.
- Dimensionality Optimization Systems: Multi-collinearity limitations in high-dimensional feature spaces; Principal Component Analysis (PCA) eigenvalue decomposition matrices; preserving total variance while reducing data dimensions.

Practical & Laboratory Component

- Feature Scaling Preparations: Implementing feature standardization (StandardScaler) and normalization (MinMaxScaler) steps prior to distance-based training.
- Cluster Engine Implementations: Running K-Means algorithms on non-indexed target data; writing automated loops to chart Elbow Curves and track shifts in distortion scores.
- Cluster Profiling Extractions: Merging generated cluster vectors back into primary data frames; extracting statistical centroids to build profiles for distinct data segments.
- PCA Dimensionality Reductions: Decomposing multi-feature spaces using PCA; charting Scree plots to check cumulative explained variance ratios across primary axes.

Unit – 5: Business Intelligence, Dashboard Design & Industrial Capstone

Theoretical Component

- Modern Business Intelligence Principles: Translating raw data metrics into live key performance indicators (KPIs); structural design choices for building scannable dashboard views.
- Data Modeling Architecture Foundations: Designing database connection structures; multi-table schema layouts (Star Schema, Snowflake Schema) and building standard cross-table filters.
- Data Analysis Expressions (DAX) Syntax Rules: Understanding evaluation context rules; calculating aggregate values, percentages, and time-intelligence calculations dynamically.
- Responsible AI Frameworks: Evaluating bias within automated decision systems; fairness considerations across demographic groups; individual data privacy standards and corporate governance metrics.

Practical & Laboratory Component

- Power BI Environment Deployments: Configuring multi-source data ingestion links; running transformation operations inside the Power Query editor workspace.
- DAX Metric Computations: Writing custom functional columns and measures to handle cross-filtering calculations and complex business evaluations.

- Interactive Canvas Construction: Assembling drill-down visual assets (dynamic cards, line tracking matrices, correlation maps) into unified corporate report pages.
- Industrial Capstone System Execution: Building an end-to-end analytics workflow using an uncurated dataset: completing data preprocessing, machine learning model runs, model checks, and a live dashboard presentation

Reference Books

1. Aggarwal, C. C. (2015). Data mining: The textbook (1st ed.). Springer.
2. Bishop, C. M. (2006). Pattern recognition and machine learning (1st ed.). Springer.
3. Downey, A. B. (2015). Think Python: How to think like a computer scientist (2nd ed.). O'Reilly Media.
4. Ferrari, A., & Russo, M. (2022). The definitive guide to DAX: Business intelligence for Microsoft Power BI, SQL Server Analysis Services, and Excel (2nd ed.). Microsoft Press.
5. Field, A. (2018). Discovering statistics using IBM SPSS Statistics (5th ed.). Sage Publications Ltd.
6. Géron, A. (2023). Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow: Concepts, tools, and techniques to build intelligent systems (3rd ed.). O'Reilly Media.
7. Grus, J. (2019). Data science from scratch: First principles with Python (2nd ed.). O'Reilly Media.
8. Hastie, T., Tibshirani, R., & Friedman, J. (2021). The elements of statistical learning: Data mining, inference, and prediction (2nd ed., corrected 12th printing). Springer.
9. James, G., Witten, D., Hastie, T., Tibshirani, R., & Taylor, J. (2023). An introduction to statistical learning: With applications in Python (1st ed.). Springer.
10. McKinney, W. (2022). Python for data analysis: Data wrangling with pandas, NumPy, and Jupyter (3rd ed.). O'Reilly Media.
11. Montgomery, D. C., & Runger, G. C. (2023). Applied statistics and probability for engineers (8th ed.). John Wiley & Sons.
12. Murphy, K. P. (2012). Machine learning: A probabilistic perspective (1st ed.). MIT Press.
13. O'Neil, C., & Schutt, R. (2013). Doing data science: Straight talk from the frontline (1st ed.). O'Reilly Media.
14. Powell, B. (2024). Mastering Microsoft Power BI (3rd ed.). Packt Publishing.
15. Provost, F., & Fawcett, T. (2013). Data science for business: What you need to know about data mining and data-analytic thinking (1st ed.). O'Reilly Media.
16. Rad, R. (2021). Pro Power BI architecture: Sharing, security, and deployment for Power BI implementations (2nd ed.). Apress.
17. Ross, S. M. (2021). Introduction to probability and statistics for engineers and scientists (7th ed.). Academic Press.

Boozy

	18. Russo, M., & Ferrari, A. (2019). Analyzing data with Power BI and Power Pivot for Excel (1st ed.). Microsoft Press. 19. Severance, C. R. (2016). Python for everybody: Exploring data using Python 3 (1st ed.). CreateSpace Independent Publishing Platform. 20. VanderPlas, J. (2016). Python data science handbook: Essential tools for working with data (1st ed.). O'Reilly Media. 21. Verzani, J. (2014). Using R for introductory statistics (2nd ed.). Chapman & Hall/CRC. 22. Witten, I. H., Frank, E., Hall, M. A., & Pal, C. J. (2017). Data mining: Practical machine learning tools and techniques (4th ed.). Morgan Kaufmann. 23. Zumel, N., & Mount, J. (2019). Practical data science with R (2nd ed.). Manning Publications
Teaching Methodology	Class/Lab teaching, Discussion, presentation, assignments, capstone project, mini projects, Case studies, Practical
Evaluation Method	Internal Evaluation: 10 (Theory)+10(practical)+10(Project) External Evaluation : 20(Theory) +30(Practical)+20(Project)

Raymond J.