**Centre for High Energy Physics**
**Faculty of Science**
**University of the Punjab, Lahore**
**Course Outline**

| **Program** | BSCP | **Course Code** | ACS 303 | **Credit Hours** | 3 (2+1Lab) |
|---|---|---|---|---|---|
| **Course Title** | **Data Science** | | | | |

## Course Introduction

In this course, we set out on a journey that makes use of the fundamental tenets of physics to unleash the enormous potential of computers and data analysis. The capacity to draw meaningful conclusions from complicated datasets is a talent of utmost significance in today's data-driven environment. As a future physicist, you will leave this course with the skills and knowledge needed to fully utilize data science. Together, we will investigate how physics may be used to describe intricate physical systems, analyze experimental data, and resolve practical issues. By the end of the course, you'll be proficient in data science approaches and have a stronger grasp of how data science may improve your comprehension and use of physics. So, let's start this thrilling trip where data and physics merge to reveal the secrets of the cosmos.

## Learning Outcomes

On the completion of the course, the students will:
This course is designed to introduce students to the basics of data science. Students will learn the fundamental concepts of data analysis, data manipulation, and data visualization using Python/R/Sql.

1. Understand the basics of data science.
2. Learn how to manipulate data in Python/R/Sql.
3. Learn how to perform basic data analysis in Python/R/Sql.
4. Lear how to visualize data in Python/R/Sql.

## Course Content

| | |
|---|---|
| **Week 1** | Introduction to Data science, Definition of data science, Applications of data science. |
| | Types of data, Data collection methods, Data cleaning, Data integration and transformation. |
| **Week 2** | Creating spreadsheets in Excel and Jupyter notebooks in python, |
| | Storing data in excel spreadsheets and Jupyter notebooks |
| **Week 3** | Using Python libraries Pandas and Numpy etc. |
| | Using Python libraries Scipy, Matplotlib, PyTorch/Tensorflow etc. |
| **Week 4** | Exploratory data analysis, Descriptive statistics.<br>• Measures of Central Tendency: Mean, Median, Mode.<br>• Measures of Dispersion: Range, Variance, Standard Deviation, Interquartile Range (IQR).<br>• Shape of the Distribution: Skewness, Kurtosis. |

| | |
|---|---|
| | Exploratory data analysis techniques.<br>Data Visualization<br>• **Histograms**: To understand the distribution of a single continuous variable.<br>• **Box Plots**: To identify outliers and understand the spread of the data.<br>• **Bar Charts**: For categorical data to show the frequency of categories.<br>• **Scatter Plots**: To examine the relationship between two continuous variables.<br>• **Pair Plots**: To see pairwise relationships between several continuous variables.<br>• **Line Plots**: Useful for time series data to show trends over time.<br>• **Heatmaps**: To show the correlation between different variables.<br>Summary Tables<br>• **Frequency Tables**: To show the frequency of categories in categorical variables.<br>• **Cross-tabulations (Contingency Tables)**: To explore the relationship between two categorical variables. |
| **Week 5** | Exploratory data analysis techniques I<br>Data Cleaning and Transformation<br>• **Missing Value Analysis**: Identifying and dealing with missing values.<br>• **Outlier Detection**: Identifying and treating outliers.<br>• **Feature Engineering**: Creating new features from existing data to improve model performance.<br>Advanced Techniques<br>• **PCA (Principal Component Analysis)**: To reduce the dimensionality of the data.<br>• **Cluster Analysis**: To identify natural groupings in the data. |
| | Exploratory data analysis techniques II<br>Bivariate and Multivariate Analysis<br>• **Correlation Matrix**: To understand the correlation between multiple variables.<br>• **Multivariate Plots**: Like pair plots, bubble plots, and 3D plots. |
| **Week 6** | Exploratory data analysis techniques III<br>**T-tests, Chi-square tests, ANOVA:** To test relationships and differences between variables. |
| | Statistical inference: Probability theory, |
| **Week 7** | Hypothesis testing confidence intervals. |
| | Regression analysis. |
| **Week 8** | Regression analysis<br>• Ohm's Law: Determine the relationship between current (I) and voltage (V) in a resistor.<br>• Projectile Motion: Model the trajectory of a projectile.<br>• Radioactive Decay: Determine the decay constant of a radioactive substance. |
| | Regression analysis |

| | |
|---|---|
| | • Thermal Expansion: Determine the coefficient of linear expansion of a material.<br>• Planck's Law: Fit the spectral radiance data to Planck's radiation law.<br>• Hubble's Law: Determine the relationship between the distance of galaxies and their recessional velocity. |
| **Week 9** | Regression analysis<br>• Damped Harmonic Oscillator: Determine the damping coefficient in a damped harmonic oscillator.<br>• Spectroscopy: Analyze the absorption spectrum to identify the concentration of a substance |
| | Regression analysis<br>• Seismic Data Analysis: Model the relationship between seismic wave velocity and depth.<br>• Gravitational Wave Analysis: Analyze the signal of gravitational waves to estimate the properties of the source. |
| **Week 10** | Introduction to big data |
| | Distributed computing |
| **Week 11** | Distributed computing I |
| | Cloud computing platforms |
| **Week 12** | Cloud computing platforms I |
| | Introduction to Hadoop and Spark. |
| **Week 13** | Data Ethics, Privacy and Security, Fairness and Bias, |
| | Data wrangling and mining from spreadsheets by using Excel<br>Accessing data from Jupyter notebooks by using Python |
| **Week 14** | APIs and web servers for data analysis |
| | Data visualization in software such as excel and Python/R/Sql |
| **Week 15** | Data analysis/visualization using CERN ROOT |
| | Histograms, Ttree, Parameter fitting etc. in ROOT |
| **Week 16** | Data Analysis using ANN.<br>Minist data sets. |
| | Data Analysis Using ANN Example |
| **Textbooks and Reading Material** | |
| 1. Python Data Science Handbook by Jake VanderPlas, O'Reilly Media Inc. (2016).<br>2. Data Science from Scratch by Joel Grus, 2nd Edition, O'Reilly Media Inc. (2018).<br>3. An Introduction to Statistical Learning with Applicatios in R/Python by Gareth James, Daniela Witten, Trevor Hastie, and Robert Tibshirani, Springer (2013).<br>4. Doing Data Science: Straight talk from frontlines by Cathy O'Neil and Rachel Schutt, O'Reilly Media, Inc.(2013).<br>5. R for Data Science by Hadley Wichkham and Garrett Grolemund, O'Reilly Media Inc (2017) | |
| **Teaching Learning Strategies** | |

The instructor is required to make use of FORTRAN/C/C++/Mathematica/Python/C# to teach the concepts through visualization/antimutation and symbolic/numerical calculations. The students are required to solve a large portion of related exercises/questions/problems of the main textbooks.

**Assignments: Types and Number with Calendar**

At least two assignments and two quizzes. A course project may also be assigned.

**Assessment**

| Sr. No. | Elements | Weightage | Details |
|---|---|---|---|
| 1. | Midterm Assessment | 35% | Written Assessment at the mid-point of the semester. |
| 2. | Formative Assessment | 25% | Continuous assessment includes Classroom participation, assignments, presentations, viva voce, attitude and behavior, hands-on-activities, short tests, projects, practical, reflections, readings, quizzes etc. |
| 3. | Final Assessment | 40% | Written Examination at the end of the semester. At least fifty percent of the question paper would involve new problems related to the concepts learned in the course.<br>It is mostly in the form of a test, but owing to the nature of the course the teacher may assess their students based on term paper, research proposal development, field work and report writing etc. |