ST. JOSEPH'S COLLEGE OF COMMERCE

(Autonomous)

163, Brigade Road, Bangalore – 560 025

Accredited with 'A++' Grade (4th Cycle) by the
National Assessment and Accreditation Council (NAAC)

Recognized by the UGC as
"COLLEGE WITH POTENTIAL FOR EXCELLENCE"

Post Graduate Diploma in Enterprise Resource Planning (Data Science)

Semester I & II

Academic year 2024-25

ST. JOSEPH'S COLLEGE OF COMMERCE

(Autonomous)

Affiliated to Bengaluru City University

St. Joseph's College of Commerce (SJCC) was formerly a part of St. Joseph's College, established in the year 1882. The Commerce Department was established in the year 1949 and it became an independent college with its own building in Brigade Road in the year 1972.

The college has in its Vision a model for higher education which encourages individuals to dream of a socially just world and in its Mission a strategy to empower individuals in realizing that dream.

With an objective of imparting quality education in the field of Commerce and Management the college has been innovating in all aspects of higher education over a long period of time. These innovations were further bolstered with the granting of autonomous status to the college by UGC in September 2005. From then on, the college has taken a lead in reforming curriculum and syllabus, examination and evaluation pattern and teaching and learning methods through the Board of Studies, the Academic Council and the Governing Council comprising of eminent academicians, industry representatives and notable alumni.

The college has undergone four cycles of NAAC accreditation starting from the year 2000 in which it secured 'five stars', next in the year 2007 an 'A' grade, in the year 2012 again an 'A' grade and recently in February 2021 an 'A++'. It is one of the very few institutions in the country to have secured A++ grade in the fourth cycle under the Revised Accreditation Framework (RAF) and the first college in Karnataka to do so. The college was declared as a 'College with Potential for Excellence' in the year 2010. In 2011 SJCC was recognized as a Research Centre by Bangalore University. The college has been ranked 93rd in the National Institutional Ranking Framework (NIRF) ratings of Ministry of Education, Government of India, in 2022 and it has been the only institution from Karnataka to make it consistently to the top 100 in the country.

OBJECTIVES

- Designed to expose learners with knowledge, skills and integrated perception of various functions of management to sustain the growing momentum of the industry and help achieve higher levels
- To provide an opportunity to students, wishing to change disciplines
- To upgrade knowledge within a discipline.
- Preparation for entry to a Master's course.

I. ELIGIBILITY FOR ADMISSION:

Candidates who have completed Graduate Programme of any recognized university and have secured passed with at least 50% of mark in the aggregate of all core papers/courses studied in the qualifying examinations are eligible for admission into this programme. Admission will be based at the discretion of the Admission Committee.

II. DURATION OF THE PROGRAMME:

The programme of the study is One (1) year of Two (2) semesters. A candidate shall complete his/her degree within Two (2) academic years from the date of his/her admission to the first semester.

III.MEDIUM OF INSTRUCTION:
The medium of instruction shall be English.

IV. ATTENDANCE:
A student shall be considered to have satisfied the requirement of attendance for the semester, if he/she has attended not less than 75% in aggregate of the number of working periods in each of the courses compulsorily. A student who fails to complete the PROGRAMME in the manner stated above shall not be permitted to take the end semester examination.

V. TEACHING AND EVALUATION:
M.Com/MBA/MFA/MBS/Ph.D graduates with B.Com/BBA/BBS as basic degree from a recognized university with a relevant industry experience are only eligible to teach and evaluate the courses.

VI. EVALUATION SYSTEM:
Evaluation for PG programme consists of two components, viz. Continuous Internal Assessment (CIA) and End Semester Examination (ESE) with the weightage of 50% and 50% respectively.

Continuous Internal Assessment (CIA) includes 20 Marks from assignments, 20 Marks from Mid-Term Test and 10 Marks from attendance. 20 Marks Assignments will be gathered from a minimum of two assignments ascertained through the exercises administered by the teacher such as Surprise test / quiz / business case analysis/ Assignment / Presentation/ Research Project/ Research article/ Seminar etc. 20 Marks will be gathered by the student through the centrally conducted Mid-Term Test by the Controller of Examination's Office, related nitty-gritties such as date, time and venue shall be notified by the department. The 10 Marks for Attendance will be calculated in the following manner - 96 – 100% = 10 Marks; 91–95% = 8 Marks; 86–90% = 6 Marks; 81-85% = 4 Marks, 76-80% = 2 Marks and up to 75% = 0 Marks. Each teaching faculty is required to maintain a record of the Continuous Internal Assessment (CIA). Under the PGD programme, a student must score a minimum of 40% (i.e., 20 Marks out of 50 Marks), in the CIA, besides 75% attendance, to be eligible for End Semester Examination, which is 12 marks through CIA.

The End Semester Examination will be conducted at the end of each semester. The duration and maximum marks for the End Semester Examination is 3 hours and for 100 marks.

VII. MINIMUM FOR A PASS:
A PGD student has to get a minimum of 40% marks in the ESE (40 on 100) and 50% aggregate in CIA & ESE (50 on 100) for a pass in each course.

VIII. PATTERN OF QUESTION PAPER:
Question Paper Pattern: (3 Hours duration, Max. Marks: 100)

Section A	Analytical questions	5 Marks x 7 Questions	35 Marks
Section B	Essay questions	15 Marks x 3 Questions	45 Marks
Section C	Compulsory questions / Case Study	20 Marks x 1 Question	20 Marks
Total Marks			**100 Marks**

PROGRAMME STRUCTURE – PGDERP (DATA SCIENCE) (2024-25)

I SEMESTER

Paper No.	Title of the Paper	Marks		Total Marks	Credits	Hours
		CIA	ESE			
P3 24 DS101	Python Fundamentals	50	50	100	4	30
P3 24 DS102	Introduction to ERP & SQL	50	50	100	4	30
P3 24 DS103	Python for Data Science	50	50	100	4	30
P3 23 DS104	Excel for Data Science	50	50	100	4	30

P3 24 DS101: PYTHON FUNDAMENTALS

Module 1: Introduction to Python **3 Hrs.**

Basics of Programming – Meaning of Python – Install Anaconda and Python – Launching a Jupyter notebook – Important Python Libraries

Module 2: Data Types & Data Structures **6 Hrs.**

Numeric – Strings – Sequence: Lists – Tuples – Sets – Mapping: Dictionaries – Arrays – Boolean

Module 3: Python Programming **6 Hrs.**

Python Programming Fundamentals: Conditions and Branching – Loops – Functions – Objects – Classes

Module 4: Working with Data in Python **6 Hrs.**

Understanding Data Processing – Python: Operations on Numpy Arrays – Overview of Data Cleaning – Slicing, Indexing, Manipulating and Cleaning

Module 5: Pandas Data frame **6 Hrs.**

Creating a Pandas Data frame – Importing data into a Pandas Data frame – Working with data using Pandas library – Handling data in Pandas: Summarizing – Aggregating and Grouping data

Module 6: Data Mining using Python **3 Hrs.**

Connecting to Database using Python – Data mining using pyodbc

BOOKS FOR REFERENCE:

1. *Fundamentals of Python: First Programs with MindTap, 2nd Edition, Kenneth A. Lambert, Cengage Learning India Pvt. Ltd., 2019*
2. Core Python Programming, 3rd Edition, R. Nageswara Rao, Dreamtech Press, 2021
3. *Fundamentals of Python, Neenu Kuriakose, Notion Press, 2022*
4. Fundamentals of Python Programming, R. Manivannan, Y. V. S. Sai pragathi, Scientific International Publishing House, 2024

P3 24 DS102: INTRODUCTION TO ERP & SQL

Module 1: Overview of ERP **3 Hrs.**

Overview of ERP, Integration of functional modules (HR/Finance/Marketing/SCM/Production), Types of data generated – Data collection, storage, extraction, cleaning – Architecture of the ERP (Application Server, Presentation Server, Database Server) – Integration & ETL Framework

Module 2: Introduction to SQL and Database **3 Hrs.**

Overview of Databases, Tables – Connecting to a Database, ODBC/ JDBC connections – connecting with excel, DB visualizer

Module 3: SQL Statement Fundamentals **6 Hrs.**

Select statement – Select Distinct – Count – Select Where – Order By – Limit – Between, IN, LIKE

Module 4: Aggregation Functions **3 Hrs.**

Group By – Having

Module 5: Join Statements **3 Hrs.**

Overview & Introduction to Joins – Inner Join – Outer Join – Left Join – Right Join – Union

Module 6: Creating Tables **6 Hrs.**

Data Types – Primary & Foreign Keys – Constraints – Create Table – Edit Table: Insert, Update, Delete, Alter, Drop

BOOKS FOR REFERENCE:

1. *SQL Database Programming Hardcover, Tim Warren, Ingram Publishing, 2020*
2. *Business Database Technology (2nd Edition): Theories and Design Process of Relational Databases, SQL, Introduction to OLAP, Overview of NoSQL Databases, Shouhong Wang, Hai Wang , Universal Publishers, 2022*
3. Introduction to DBMS, Sana Shaikh, BPB Publishing House, 2022
4. *SQL Primer: An Accelerated Introduction to SQL Basics, Rahul Batra, Apress, 2018*
5. Bek's Introduction to SQL Programming: "That Pelican Database Book", Jon Bek, 2024

P3 24 DS103: PYTHON FOR DATA SCIENCE

Module 1: Explore Data using Python **6 Hrs.**

Understanding Characteristics of the dataset – Size, Shape, value counts – Data Types and change data types

Module 2: Measures of Central Tendency **5 Hrs.**

Arithmetic Mean, Median, Mode – Relationship between mean, median and mode – Computation of the measures for grouped and ungrouped data

Module 3: Measures of dispersion **6 Hrs.**

Range, mean deviation and standard deviation – coefficient of variation and its use – Quartiles and Inter quartile range – Quintiles, Deciles and Percentiles — Skewness and Kurtosis and their uses

Module 4: Understanding Correlation **3 Hrs.**

Understanding correlation – calculating correlation using Pandas – Interpreting a correlation matrix

Module 5: Statistical Quality Control **3 Hrs.**

Nature of Control Limits – Purpose of Control Charts – Control Charts for Variables – Control Charts for Attributes

Module 6: Conducting EDA using Python **6 Hrs.**

Data Analysis using Python – Handling missing data – Computing metrics – Analysis & Interpretation of the data and connected visualization

BOOKS FOR REFERENCE:

1. *Hands-On Data Science and Python Machine Learning, Kane Frank, Packt Publishing Limited*
2. *Data Science and Machine Learning Interview Questions Using Python a Complete Question Bank to Crack Your Interview, Vishwanathan Narayanan, BPB Publications.*
3. *Python for data science, Tony F Charles.*
4. *Data Structures using Python, Shriram K. Vasudevan, OUP India.*
5. *Python Data Science Handbook, Jake Vanderplas, O'Reilly Media.*

P3 23 DS104: EXCEL FOR DATA SCIENCE

COURSE OBJECTIVES:

The students should be able to

1. Explain the Role of Spreadsheets in any business.
2. Design a Dashboard for the business to measure business goals
3. Calculate various metrics using formulas & functions
4. Evaluate data provided using data handling tools and analysis packages
5. Examine the data using charts & graphs

Module 1: Introduction to Spreadsheets **3 Hrs**

Spreadsheet tools Excel and Google Sheets – fundamentals of spreadsheet applications – Data entry – Data gathering – Working with Microsoft Excel

Module 2: Excel Formulas **6 Hrs**

Excel Formulas – basic & advanced: Mathematical Formulas – Logical – Text – Date & Time– Look-up

Module 3: Data handling in Excel **5 Hrs**

Data types, Data structures, Data handling in Excel: Data Tools – Filtering – Sorting – Group & Ungroup – Conditional Formatting – Data Validation

Module 4: Working with Pivot tables **5 Hrs**

Working with Pivot tables: Meaning of Pivot table – Create a Pivot table – Data Source & Layouts – Calculation & Formatting – Sort, Filter & Extract – Data Visualization

Module 5: Exploratory data Analysis in Excel **7 Hrs**

Understand Characteristics of the dataset – Basic Statistics in Excel using Analysis Toolpak Add-in – Measure of central tendency (Mean, Median, Mode) – Measure of spread (Range, Quartile, Percentiles, absolute deviation, variance and standard deviation) – Measure of symmetry (Skewness) – Measure of Peakedness (Kurtosis)

Module 6: Visualization & Dashboards in Excel **4 Hrs**

Charts in Excel: Show the data – Formatting – Creating Simple Dashboards in Excel – Static & Dynamic dashboards

BOOKS FOR REFERENCE:

1. Excel Data Analysis, Hector Guerrero, Springer - Verlag Berlin and Heidelberg GmbH & Co. KG.
2. How to Excel In Data Science Interview, Ken J Stevens, Createspace Independent Publishing Platform.
3. Data, Statistics, and Decision Models with Excel, Donald L. Harnett and James F. Horrell, John Wiley & Sons Inc.
4. EXCEL FOR DATA SCIENCE, Dr Bienbenue Maula
5. Automated Data Analysis Using Excel, Brian D. Bissett, Apple Academic Press Inc.

PROGRAMME STRUCTURE – PGDERP (DATA SCIENCE) (2024-25)

II SEMESTER

Paper No.	Title of the Paper	Marks		Total Marks	Credits	Hours
		CIA	ESE			
P3 23 DS201	Advanced Statistical Methods - I	50	50	100	4	30
P3 23 DS202	Advanced Statistical Methods - II	50	50	100	4	30
P3 24 DSMH203	Data Science Applications – Marketing, Sales and HR	50	50	100	4	30
P324DSBA204	Data Visualization and Dashboard Design	50	50	100	4	30

P3 23 DS201: ADVANCED STATISTICAL METHODS - I

COURSE OBJECTIVES:

The students should be able to

1. Explain the Role of Financial Manager in the present state of Affairs.

2. Design an optimum Capital structure that minimises the overall Cost of Capital

3. Calculate Risk and Return of portfolio by using CAPM.

4. Evaluate Investment Decision by using Capital Budgeting Technique under Risk and Uncertainty.

5. Examine the extent to which Capital Markets are efficient and its implications on the role of Capital Market in Merchant Banking.

Module 1: Linear Regression **4 Hrs**

Linear Regression – the theory: What is simple linear regression and its uses – Types of variables – Assumptions of simple linear regression – Perform a simple linear regression involving test & training data sets – Interpreting the results – Presenting the results.

Module 2: Linear Regression using Python **6 Hrs**

Carry out Simple Linear regression analysis in Python and interpreting the results using pandas, numpy, maplotlib, seaborn, sklearn.

Module 3: Multi Linear Regression **4 Hrs**

Multi Linear Regression – the theory: What is Multi linear regression and its uses – Types of variables – Assumptions of Multi linear regression – Perform a Multi linear regression – Interpreting the results – Presenting the results

Module 4: Multi Linear Regression using Python **6 Hrs**

Carry out Multi-Linear regression analysis in Python and interpreting the results using pandas, numpy, maplotlib, seaborn, sklearn

Module 5: Logistic Regression **4 Hrs**

Logistic Regression – the theory: What is Logistic regression and its uses – Linear regression vs logistic regression – Types of variables – Assumptions of Logistic regression – Types of Logistic regression models – Perform a Logistic regression – Interpreting the results – Presenting the results

Module 6: Logistic Regression using Python **6 Hrs**

Carry out Binary Logistic regression analysis in Python and interpreting the results using pandas, numpy, maplotlib, seaborn, sklearn

BOOKS FOR REFERENCE:

1. Statistical Methods by SP GUPTA 48TH EDITION 2022, Sultan Chand & Sons.

2. Advanced Statistical Methods in Data Science, Xuewen Lu, Ding-Geng Chen, Jiahua Chen, Hao Yu , Grace Y. Yi, Springer

3. Advanced Statistical Methods for the Analysis of Large Data-Sets, Agostino Di Ciaccio, Mauro Coli, Jose Miguel Angulo Ibanez, Springer-Verlag Berlin and Heidelberg GmbH & Co. KG

4. Advanced and Multivariate Statistical Methods, Rachel A. Vannatta, Craig a Mertler, Craig A. (Arizona State University, USA) Mertler, Kristina N Lavenia, Rachel A Vannatta, Taylor and Francis.
5. Understanding Advanced Statistical Methods, Peter Westfall, Kevin S. S. Henning, Peter (Texas Tech University, Lubbock, USA) Westfall, Kevin S. S. (Sam Houston State University, Huntsville, Texas, USA) Henning, Kevin S S Henning, CRC Press.

P3 23 DS202: ADVANCED STATISTICAL METHODS - II

COURSE OBJECTIVES:

The students should be able to

1. Explain briefly unsupervised learning problems and algorithms used to solve them.

2. Design a Time series, Market basket and Clustering analysis to solve business problems using Python

3. Calculate various metrics involved in the Time series, Market basket and Clustering analysis

4. Evaluate business problems and suggest the right unsupervised learning technique to solve it.

5. Examine application of these techniques in different industries.

Module 1: Time Series Analysis **4 Hrs**

Introduction to Time Series Analysis – Components – Types of data – Terminology – Visualize the Time Series – Patterns in a Time Series – Additive and Multiplicative Time Series – Difference between white noise and a stationary series

Module 2: Time Series Analysis using Python **6 Hrs**

Carry out time-series analysis in Python and interpreting the results using pandas, numpy, maplotlib, seaborn

Module 3: Clustering Analysis **4 Hrs**

Introduction to Unsupervised learning problems – Introduction to Clustering – Types of Clustering & Clustering Algorithms – K means clustering – Applications of Clustering

Module 4: K means clustering using Python **6 Hrs**

Carry out K Means Algorithm in Python using scikit-learn library

Module 5: Association Rule mining **4 Hrs**

Meaning of Association rules – Algorithms used in Market Basket analysis – Advantages of Market basket analysis – Market basket analysis from Customer and Retailer perspectives

Module 6: Market Basket Analysis using Python **6 Hrs**

Carry out Market Basket Analysis from scratch in Python

BOOKS FOR REFERENCE:

1. Statistical Methods by SP GUPTA 48TH EDITION 2022, Sultan Chand & Sons.

2. Advanced Statistical Methods in Data Science, Xuewen Lu, Ding-Geng Chen, Jiahua Chen, Hao Yu , Grace Y. Yi, Springer

3. Advanced Statistical Methods for the Analysis of Large Data-Sets, Agostino Di Ciaccio, Mauro Coli, Jose Miguel Angulo Ibanez, Springer-Verlag Berlin and Heidelberg GmbH & Co. KG

4. Advanced and Multivariate Statistical Methods, Rachel A. Vannatta, Craig a Mertler, Craig A. (Arizona State University, USA) Mertler, Kristina N Lavenia, Rachel A Vannatta, Taylor and Francis.

5. Understanding Advanced Statistical Methods, Peter Westfall, Kevin S. S. Henning, Peter (Texas Tech University, Lubbock, USA) Westfall, Kevin S. S. (Sam Houston State University, Huntsville, Texas, USA) Henning, Kevin S S Henning, CRC Press.

P3 24 DSMH203: DATA SCIENCE APPLICATIONS –MARKETING, SALES AND HR

Module 1: Data Science Application in Marketing, Sales and HR 3 hrs.
Need for Analytics – Types of Analytics used in Marketing, Sales and HR – Measuring benefit of data driven initiatives

Module 2: Data Types generated in Marketing & Sales 3 Hrs.
Understand the various data generated in Marketing, Sales and HR departments in the business – Primary data & Secondary data

Module 3: Retail Analytics & Sales 5 Hrs.
Trends & growth – Sales target & achievement – Lead to Conversion analysis – Pipeline analysis – Segmentation – Sales Cycle – Analyzing Sales & Retail data

Module 4: Customer Analytics 5 Hrs.
Customer Segmentation – RFM analysis using Excel – Customer churn – Understanding Customer loyalty using CLI model

Module 5: Types of HR Metrics 5 Hrs.
Understand the strategic KPI used in HR – HR metrics: Recruitment – Learning & Development – Operations – Compensation & Benefits – Employee Engagement – Time tracking – Employee Performance & Rewards

Module 6: Human Resource Analytics 5 Hrs.
Employee turnover – Attrition analytics – Measuring Employee engagement & satisfaction – Key drivers of Employee engagement – Diversity KPIs

BOOKS FOR REFERENCE:

1. *Marketing Data Science: Modeling Techniques in Predictive Analytics with R and Python (FT Press Analytics) Thomas W. Miller, Pearson.*
2. *Practical Text Analytics: Interpreting Text and Unstructured Data for Business Intelligence (Marketing Science), Dr. Steven Struhl,* Kogan Page.
3. *Data Science for Marketing Analytics, Mirza Rahim Baig, Gururajan Govindan, Packt Publishing Limited.*
4. *Big Data Impacts on Human Resource Management. Using Multinational Enterprises as Example, Yi-Chuan Chen, Grin Publishing.*
5. *Big Data in Organizations and the Role of Human Resource Management, Tobias M. Scholz, Peter Lang AG.*
6. *Application of fuzzy techniques to Human Resources Management, Trinidad Casasus Estelles, LAP Lambert Academic Publishing.*

P324DSBA204: DATA VISUALIZATION & DASHBOARD DESIGN

Module 1: Introduction to Visualization **3 hrs.**

Need to Visualize data – Visualization as an essential tool for communication – Seven stages of data visualization – Exploratory and Explanatory analysis

Module 2: Components of a Data Visualization **6 hrs.**

Different types of Data – Different types of Graphs – Right graph for the right data – Bringing out the Story with Color, labeling and formatting – Static vs. Interactive Visualization

Module 3: Dash-boarding Fundamentals **3 Hrs.**

Meaning of Dashboard – Uses – Roles & Messages: author-driven vs. reader-driven – Different types of Dashboards – Dash-boarding tools – Install Microsoft Power-BI and its overview

Module 4: Building an Operational Dashboard using Power BI **3 Hrs.**

Meaning of KPI – Data required for a KPI/ Operational Dashboard – KPI dashboard use cases and associated metrics – Building an interactive KPI Dashboard in Power BI

Module 5: Building an Analytical Dashboard using Power BI **3 Hrs.**

Meaning of Analytical Dashboard – Data required for the Dashboard – Building an interactive Analytical Dashboard in Power BI

Module 6: Building a Strategic Dashboard using Power BI **3 Hrs.**

Meaning of Strategic Dashboard – Data required for a Strategic/ Management Dashboard – Building an interactive Strategic Dashboard using Power BI

BOOKS FOR REFERENCE:

1. *STORYTELLING WITH DATA Paperback – Illustrated, Cole Nussbaumer Knaflic, Wiley, 2019*
2. *Data Visualization Made Simple, Kristen Sosulski, Taylor and Francis Ltd, 1st Edition 2021.*
3. *Data Visualization for Design Thinking, Newman, Winifred, Taylor and Francis, 2022.*
4. *Data Storytelling and Visualization with Tableau, Hardbound by Joshi, Prachi Manoj, Taylor and Francis Ltd, 1st Edition 2022.*