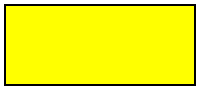
SRI A.S.N.M GOVERNEMENT COLLEGE (A), PALAKOL

B.Sc (Computer Science)

(For Data Science)

Syllabus and Model Papers

Employability 

Entrepreneurship 

Skill Development 

SRI A.S.N.M. GOVERNMENT COLLEGE (A), PALAKOL

CBCS/Semester System
(W.e.f. 2020-21 Admitted Batch)

DEPARTMENT OF COMPUTER SCIENCE & APPLICATIONS

B.Sc (Data Science) Structure of Syllabus

Sem		Course Code	Course Name	Total Marks	Max. Marks Cont/ Internal /Mid Assessment	Max. Marks Sem-end Exam	Hrs/ Week	Credits (3+2)
FIRST YEAR								
I SEM			INTRODUCTION TO DATA SCIENCEAND R PROGRAMMING	100	**40**	**60**	4	3
			INTRODUCTION TO DATA SCIENCEAND R PROGRAMMING LAB	50	**0**	**50**	2	2
II SEM			DATA MINING CONCEPTS AND TECHNIQUES	100	**40**	**60**	4	3
			DATA MINING CONCEPTS AND TECHNIQUES LAB	50	**0**	**50**	2	2
SECOND YEAR								
III SEM			PYTHON PROGRAMMIN GFOR DATA ANALYSIS	100	25	**75**	4	3
			PYTHON PROGRAMMIN GFOR DATA ANALYSIS LAB	50	0	**50**	2	2
IV SEM			BIG DATA ANALYTICS USINGSPARK	100	25	**75**	4	3
			BIG DATA ANALYTICS USINGSPARK LAB	50	0	**50**	2	2
			DATA VISUALIZATION	100	25	**75**	4	3
			DATA VISUALIZATION LAB	50	0	**50**	2	2

SRI A.S.N.M. GOVERNMENT COLLEGE (A), PALAKOL

CBCS/Semester System
(W.e.f. 2020-21 Admitted Batch)

B.Sc (Data Science)
I YEAR I SEMESTER SYLLABUS

Introduction to Data Science and R Programming

Unit-1

Defining, Data Science and Big data, Benefits and Uses, facets of Data, Data Science Process. History and Overview of R, Getting Started with R, R Nuts and Bolts.

Unit-2

The Data Science Process: Overview of the Data Science Process-Setting the research goal, Retrieving Data, Data Preparation, Exploration, Modeling, data Presentation and Automation. Getting Data in and out of R, Using reader package, Interfaces to the outside world.

Unit-3

Machine Learning: Understanding why data scientists use machine learning-What is machine learning and why we should care about, Applications of machine learning in data science, Where it is used in data science, The modeling process, Types of Machine Learning-Supervised and Unsupervised.

Unit-4

Handling large Data on a Single Computer: The problems we face when handling large data, General Techniques for handling large volumes of data, Generating programming tips for dealing with large datasets. Case study- Predicting malicious URLs (This can be implemented in R).

Unit-5

Sub setting R objects, Vectorised Operations, Managing Data Frames with the dplyr, Control structures, and functions, Scoping rules of R, Coding Standards in R, Loop Functions, Debugging, and Simulation.

Recommended Text books:

1. DavyCielen, Arno.D.B.Maysman, Mohamed Ali, “Introducing Data Science” Manning Publications, 2016.
2. Roger D. Peng, “R Programming for DataScience” Lean Publishing, 2015.
3. Nina Zumel, John Mount, “Practical Data Science with R”, Manning Publications, 2014.
4. Tony Ojeda, Sean Patrick Murphy, Benjamin Bengfort, AbhijitDasgupta, “Practical Data Science Cookbook”, Packt Publishing Ltd., 2014.

SRI A.S.N.M. GOVERNMENT COLLEGE (A), PALAKOL

CBCS/Semester System
(W.e.f. 2020-21 Admitted Batch)

B.Sc (Data Science)
I YEAR I SEMESTER SYLLABUS

Introduction to Data Science and R Programming Lab

1) Installing R and R studio
2) Basic operations in r
3) Getting data into R, Basic data manipulation, Loading Data into R
4) Basic plotting
5) Loops and functions
6) Create Vectors, Lists, Arrays, Matrices, Data frames and operations on them.
7) Demonstrate the visualization and graphics using visualization packages.
8) Implement Loop functions with lapply(), sapply(), tapply(), apply(), mapply().
9) Explore data using Single Variables: Unimodal, Bimodal, Histograms, Density Plots, Bar charts
10) Explore data using two Variables: Line plots, Scatter Plots, smoothing cures, Bar charts
11) Explore and implement commands using dplyr package
12) Generate random numbers and set seed

Recommended Text books:

Mark Gardener, "Beginning R - The Statistical Programming Language", John Wiley & Sons, Inc., 2012.

W. N. Venables, D. M. Smith and the R Core Team, "An Introduction to R", 2013.

Recommended Reference books:

1. The art of R Programming: A tour of Statistical Software design. Norman Matloff. Kindle Edition

2. The book of R : The first course in Programming and Statistics by Tilman M. Davies.

SRI A.S.N.M. GOVERNMENT COLLEGE (A), PALAKOL

CBCS/Semester System
(W.e.f. 2020-21 Admitted Batch)
B.Sc (Data Science)

I YEAR I SEMESTER

Introduction to Data Science and R Programming
Model Question Paper

Max. Marks: 75 **Time: 3Hrs Max**

SECTION-A **5x5=25M**
(Answer any Five of the following)

1. What is data science, and big data, how data science and Big data are related. What is the application of data science?
2. Explain Read R package
3. What are the applications of machine learning in data science.
4. What are the different challenges that w face when handling large data.
5. What is meant by data frame in 'R'? Explain dplyr package.
6. What are the different types of big data?
7. What are the four steps in modeling process in machine earning?
8. What is meant by debugging?

SECTION-B **5x10=50M**

9. (a) Explain different phases of facets of data.

(OR)

(b) What is R. Describe basic commends in R with Examples (Vectors, matrices, lists, data frames etc.)

10. (a) Explaining detail the steps involved in data science process.

(OR)

(b) What are the different ways of leading data into R? Explain with examples.

11. (a) What are the different types of machine learning processes? Explain detail.

(OR)

(b) List out the importance of machine learning and gives examples in our day to day life.

12. (a) What are the different techniques for handling large volumes of data?

(OR)

(b) Explain any case study that deals with large data sets.

13. (a) Explain Vectorised operations, control structures, functions and loop functions in R.

(OR)

(b) b) Explain and give examples of exploring data using single variable and two variables.

SRI A.S.N.M. GOVERNMENT COLLEGE (A), PALAKOL

CBCS/Semester System
(W.e.f. 2020-21 Admitted Batch)
B.Sc (Data Science)

I YEAR II SEMESTER SYLLABUS

DATA MINING CONCEPTS AND TECHNIQUES

Unit-I

An idea on Data Warehouse, Data mining-KDD versus data mining, Stages of the Data Mining Process-Task primitives. Data Mining Techniques – Data mining knowledge representation.

Unit-II

Data mining query languages- Integration of Data Mining System with a Data Warehouse-Issues, Data pre-processing – Data Cleaning, Data transformation – Feature selection – Dimensionality reduction

Unit-III

Concept Description: Characterization and comparison What is Concept Description,Data Generalization by Attribute-Oriented Induction(AOI), AOI for Data Characterization, Efficient Implementation of AOI.
Mining Frequent Patterns, Associations and Correlations: Basic Concepts, FrequentItemset Mining Methods: Apriori method, generating Association Rules, Improvingthe Efficiency of Apriori, and Pattern-Growth Approach for mining Frequent Item sets.

UNIT-IV

Classification Basic Concepts: Basic Concepts, Decision Tree Induction: Decision TreeInduction Algorithm, Attribute Selection Measures, Tree Pruning. Bayes Classification Methods.

UNIT-V

Classification by Back Propagation:Multi_Layer Feed Forward Neural Network.
Support Vector Machines: Cases when the data are linearly separable and linearly inseparable.
Cluster Analysis: Cluster Analysis, Partitioning Methods, Hierarchal methods, Density based methods-DBSCAN.

References

1. Jiawei Han and MichelineKamber, “Data Mining: Concepts and Techniques”, 3rd Edition, Morgan Kaufmann Publishers, 2011.
2. AdelchiAzzalini, Bruno Scapa, “Data Analysis and Data mining” , 2ndEdiiton, Oxford Univeristy Press Inc., 2012.
3. Alex Berson and Stephen J. Smith, “Data Warehousing, Data Mining & OLAP”, 10th Edition, TataMcGraw Hill Edition , 2007.
4. G.K. Gupta, “Introduction to Data Mining with Case Studies”, 1st Edition, Easter Economy Edition, PHI, 2006.

SRI A.S.N.M. GOVERNMENT COLLEGE (A), PALAKOL

CBCS/Semester System
(W.e.f. 2020-21 Admitted Batch)
B.Sc (Data Science)

I YEAR II SEMESTER SYLLABUS

DATA MINIG USING R PROGRAMMING LAB

1. Get and Clean data using swirl exercises.(Use 'swirl' package, library and install that topic from swirl).
2. Visualize all Statistical measures(Mean ,Mode, Median, Range, Inter Quartile Range etc., using Histograms, Boxplots and Scatter Plots).
3. Create a data frame with the following structure.

EMP ID	EMP NAME	SALARY	START DATE
1	Satish	5000	01-11-2013
2	Vani	7500	05-06-2011
3	Ramesh	10000	21-09-1999
4	Praveen	9500	13-09-2005
5	Pallavi	4500	23-10-2000

a. Extract two column names using column name.
b. Extract the first two rows and then all columns.
c. Extract 3rd and 5th row with 2nd and 4th column.

4. Create a data frame with 10 observations and 3 variables and add new rows and columns to it using 'rbind' and 'cbind' function.

5. Create a function to discretize a numeric variable into 3 quantiles and label them as low, medium, and high. Apply it on each attribute of any dataset to create a new data frame. 'discrete' with Categorical variables and the class label.

6. Create a simple scatter plot using any dataset using 'dplyr' library. Use the same data to indicate distribution densities using box whiskers.

7. Write R Programs to implement k-means clustering, k-medoids clustering and density based clustering on any datasets.

8. Write a R Program to implement decision trees using 'reading Skills' dataset.

9. Implement decision trees using any dataset using package party and 'rpart'.

10. Train SVM Model by taking any dataset.

SRI A.S.N.M. GOVERNMENT COLLEGE (A), PALAKOL

CBCS/Semester System
(W.e.f. 2020-21 Admitted Batch)
B.Sc (Data Science)

I YEAR II SEMESTER

DATA MINING CONCEPTS AND TECHNIQUES

Model Question Paper

Max. Marks: 75 **Time: 3Hrs Max**

SECTION-A **5x5=25M**

(Answer any Five of the following)

1. What is Data mining explain the architecture of Data mining.
2. Discuss issues to be considered during data integration of Data mining system with a ware house.
3. Explain Apriori method.
4. State Bayes theorem and explain Bayesian belief network.
5. Define support and confidence in association rule mining.
6. Discuss reasons to perform data pre-processing.
7. Describe data characterisation.
8. What is SVM? Explain linearly separable data.

SECTION-B **5x10=50M**

9. What is Data mining functionality? Explain different types of Data mining functionalities with examples.

   (OR)

   Discuss in detail about the steps in knowledge discovery in data bases. Explain different techniques in Data mining.
10. Describe the process of data cleaning and data transformation In pre processing

    (OR)

    Explain various data reduction and dimensionality reduction in the pre processing step of Data mining.
11. Discuss concept description and generalised by AOI for data characterisation.

    (OR)

    Frequent item set mining methods by frequent pattern mining algorithm.
12. Explain the algorithm for construction a decision tree from training samples.

    (OR)

    Explain Basian theorem.
13. Multifeed forward neural networks

    (OR)

    What is cluster? Explain how we form clusters through K-means.

SRI A.S.N.M. GOVERNMENT COLLEGE (A), PALAKOL

CBCS/Semester System
(W.e.f. 2020-21 Admitted Batch)
B.Sc (Data Science)

II YEAR III SEMESTER SYLLABUS

PYTHON PROGRAMMING FOR DATA ANALYSIS

UNIT- I

What is Data Analysis? Differences between Data Analysis and Analytics, What is Python, Why Python for Data Analysis? What is Library, Essential Python Libraries. Python Language basics, I Python and Jupyter Notebook. Python Language Basics.

UNIT- II

Built-in Data Structures, Functions, Files and Operating System.

NumPy Basics: Arrays and Vectorized Computation, The Numpynd array, Universal Functions, Array-Oriented Programming with Arrays, File Input and Output with Arrays, Linear Algebra, Pseudorandom Number Generation.

UNIT-III

Getting Started with Pandas: Introduction to Pandas Data Structures, Essential Functionality, Summarizing and Computing Descriptive Statistics

Data Loading, Storage and File Formats: Reading and Writing Data in Text Format, Binary Data Formats, Interacting with Web APIs, Interacting with Databases.

UNIT- IV

Data Cleaning and Preparation: Handling Missing Data, Data Transformation, String Manipulation.

Data Wrangling: Join, Combine and Reshape: Hierarchical Indexing, Combining and Merging Datasets, Reshaping and Pivoting.

UNIT -V

Introduction to Modeling Libraries in Python: Interfacing between pandas and Model code, Creating model descriptions with Patsy, Introduction to stats models.

Plotting and Visualization: A brief matplotlib API Primer, Plotting with Pandas and Seaborn, Other Python visualization tools.

Reference Books

1. Wes McKinney “Python for Data Analysis” O’reilly Publications Second edition
2. Charles R Suverance “Python for Everybody” Exploring data using Python 3
3. John Zelle Michael Smith Python Programming, second edition 2010

SRI A.S.N.M. GOVERNMENT COLLEGE (A), PALAKOL

CBCS/Semester System
(W.e.f. 2020-21 Admitted Batch)
B.Sc (Data Science)

II YEAR III SEMESTER

PYTHON PROGRAMMING LAB

1. Use matplotlib and plot an inline in Jupyter.
2. Implement commands of Python Language basics
3. Create Tuples, Lists and illustrate slicing conventions.
4. Create built-in sequence functions.
5. Clean the elements and transform them by using List, Set and Dict Comprehensions.
6. Create a functional pattern to modify the strings in a high level.
7. Write a Python Program to cast a string to a floating-point number but fails with Value
   Error on improper inputs using Errors and Exception handling.
8. Create an n array object and use operations on it.
9. Use arithmetic operations on Numpy Arrays
10. Using Numpy array perform Indexing and Slicing Boolean Indexing, FancyIndexing operations
11. Create an image plot from a two-dimensional array of function values.
12. Implement some basic array statistical methods (sum, mean, std, var, min,max, argmin, argmax, cumsum andcumprod) and sorting with sort method.
13. Implement numpy.random functions.
14. Plot the first 100 values on the values obtained from random walks.
15. Create a data frame using pandas and retrieve the rows and columns in it byperforming some indexing options and transpose it.
16. Implement the methods of descriptive and summary statistics
17. Load and write the data from and to different file formats including WebAPIs.
18. Implement the data Cleaning and Filtering methods(Use NA handlingmethods, fillna function arguments)
19. Transform the data using function or mapping
20. Rearrange the data using unstack method of hierarchical Indexing
21. Implement the methods that summarize the statistics by levels.
22. Use different Join types with how argument and merge data with keys and multiple keys.

SRI A.S.N.M. GOVERNMENT COLLEGE (A), PALAKOL

CBCS/Semester System
(W.e.f. 2020-21 Admitted Batch)

B.Sc (Data Science)
II YEAR III SEMESTER

Python programming for data analysis
Model Question Paper

Max. Marks: 75 **Time: 3Hrs Max**

SECTION-A **5x5=25M**
(Answer any Five of the following)

1) What is Data analysis and Data analytics? What are the differences between them?

2) Explain different built in data structures in python

3) How pandas are used in Python.

4) Explain Reshaping and pivoting.

5) What is Pandas?

6) Explain Universal functions

7) Explain interactive with data base concepts.

8) Explain different python visualization tools.

SECTION-B **5x10=50M**

9) a) Why python is used for data analysis. What is meant by library and explain at least six python libraries.

(or)

b) What are I python and Jupiter note book? Why they are used.

10) a) What is meant by numpy. Why and how numpy is used in python. Explain with in an example.

(or)

b) Write a program to generate a pseudo random number in python and write a program find out the number of elements in an array.

11) a) Explain predictive and descriptive statistics. Explain with formulas.

(or)

b) Explain how the data is loaded, stored in different file formats in python.

12) a) What are the different data cleaning and preparation methods?

Explain.

(or)

b) Write python program on hierarchical indexing and joint and combining data.

13) a) How to create model description in python. Explain with a program.

(or)

b) Matplotlib is used for plotting and visualization in python using that package explain with example.

SRI A.S.N.M. GOVERNMENT COLLEGE (A), PALAKOL

CBCS/Semester System
(W.e.f. 2020-21 Admitted Batch)
B.Sc (Data Science)

II YEAR IV SEMESTER SYLLABUS

BIG DATA ANALYTICS USING SPARK

UNIT - I

Introduction to Big Data: What is Big Data-Characteristics, Data in the Warehouse and Data in Hadoop, Why is Big Data Important- When to consider Big Data Solution, Applications.
Introduction to Hadoop: Hadoop- definition, Application development in Hadoop. The building blocks of Hadoop, Name Node, Data Node, Secondary Name Node, Job Tracker and Task Tracker.

UNIT-II

Introduction to Spark: What is Apache Spark, Why Spark when Hadoop is there, Spark Features, Spark components, Spark program flow, Spark Eco System? Differences between implementation of programs in Hadoop and Spark Programming environments.

UNIT III

Spark Fundamentals- Using spark in action VM, Using Spark Shell and writing first spark program, Basic RDD actions and transformations.
Spark SQL-Working with Data Frames, Using SQL Commands, Saving and loading Data Frame.

UNIT IV

Streaming in Spark- Writing spark streaming applications, using external data sources, structured streaming**.**
Spark ML lib-Introduction to Machine Learning. Definition of Machine Learning, Machine Learning with Spark.

UNIT V

Graph Representation in MapReduce: Graph Processing with Spark, Spark GraphX, GraphX features, Graph Examples, Graph algorithms-Shortest Path Algorithm.

REFERENCE BOOKS:

1. Understanding Big Data Analytics for Enterprise Class Hadoop and Streaming Data by Dirk deRoos, Chris Eaton, George Lapis, Paul Zikopoulos, Tom Deutsch, 1st Edition, TMH,2012.
2. Spark in Action PetarZecevic, markoBonaci Manning Publications-2016.
3. Learning Spark"Holden KarauA. Konwinskietc.,"O'reilly Publications.
4. Hadoop in Action by Chuck Lam, MANNING Publishers.
5. Hadoop: The Definitive Guide by Tom White, 3rd Edition, O'reilly
6. Mining of massive datasets, AnandRajaraman, Jeffrey D Ullman, Wiley Publications.

SRI A.S.N.M. GOVERNMENT COLLEGE (A), PALAKOL

CBCS/Semester System
(W.e.f. 2020-21 Admitted Batch)
B.Sc (Data Science)

II YEAR IV SEMESTER

SPARK PROGRAMMING LAB

1. Using Python Implement the following Programs
   a) Write Program to implement arithmetic operations
   b) Write Program to find the biggest of two numbers
   c) Write a program to find the matrix multiplication
2. Install Hadoop
3. Install Spark on top of Hadoop
4. Create and Implement the transformations in RDDs
5. Create a data frame from an existing RDD using Spark Session
6. Execute a Word Count example in Spark Shell by creating RDDs.
7. Implement Spark SQL Queries in Python.
8. Write a Program to implement maximum temperature give the recordings of one year.
9. Write a Program to implement the Pie estimation
10. Write a User Defined Function to convert a given text to Uppercase.

SRI A.S.N.M. GOVERNMENT COLLEGE (A), PALAKOL

CBCS/Semester System
(W.e.f. 2020-21 Admitted Batch)
B.Sc (Data Science)

II YEAR IV SEMESTER

Big Data Analytics using Spark
Model Question Paper

Max. Marks: 75 **Time: 3Hrs Max**

SECTION-A **5x5=25M**
(Answer any Five of the following)

1) What is big data? What are its characteristics?

2) Why we have to use spark when Hadoop is there?

3) What are the data structures in spark? Explain the concept of RDD is spark?

4) Write the applications of spark streaming

5) Explain the features of spark graphics?

6) What is meant by Hadoop define.

7) What are the differences between data frames and data sets in spark?

8) Explain the concept of machine learning?

SECTION-B **5X10=50M**

9) a) What are the differences between the data in hadoop and in warehouse

(or)

b) Explain the building blocks of hadoop

10) a) Explain the components of spark and program flow in spark?

(or)

b) Explain difference between implementation of programs in Hadoop and spark programming environment?

11) a) Explain RDD transmission and actions

(or)

b) With spark SQL commends explain how to save and load data in data frame

12) a) Explain different extend data

sources (or)

b) How to implement machine learning concept in spark?

13) a) Explain graphs processing with spark using map reduce

(or)

b) Explain shortest path algorithm

SRI A.S.N.M. GOVERNMENT COLLEGE (A), PALAKOL

CBCS/Semester System
(W.e.f. 2020-21 Admitted Batch)
B.Sc (Data Science)

II YEAR IV SEMESTER SYLLABUS

DATA VISUALIZATION

UNIT I

Creating Visual Analytics with tableau desktop, connecting to your data-How to Connect to your data, what are generated Values? Knowing when to use a direct connection, joining tables with tableau, blending different data sources in a single worksheet.

UNIT II

Building your first Visualization- How Me works- Chart types, Text Tables, Maps, bar chart, Line charts, Area Fill charts and Pie charts, scatter plot, Bullet graph, Gantt charts, Sorting data in tableau, Enhancing Views with filters, sets groups and hierarchies.

UNIT III

Creating calculations to enhance your data- What is aggregation, what are calculated values and table calculations, using the calculation dialog box to create, Building formulas using table calculations, using table calculation functions

UNIT IV

Using maps to improve insights-Create a Standard Map View, Plotting your own locations on a map, Replace Tableau's standard maps, shaping data to enable Point-to-Point mapping.

UNIT V

Developing an Adhoc analysis environment- generating new data with forecasts, providing self-evidence adhoc analysis with parameters, Editing views in tableau Server.

Reference Books

1. Tableau your data-Daniel G. Murray and the Inter works BI team, Wiley Publications
2. Tableau Data Visualizaton Cookbook, AshutoshNandeshwar, PACKT publishing.
3. Storytelling with Data: A Data Visualization Guide for Business Professionals by Cole NussbaumerKnaflic (2014)
4. ggplot2: Elegant Graphics for Data Analysis by Hadley Wickham (2009)
5. Designing Data Visualizations: Representing Informational Relationships by Noah Iliinsky, Julie Steele (2011)
6. Alexandru C. Telea – "Data Visualization principles and practice" Second Edition, CRC Publications

Joshua N. Millign–" Learning Tableau -2019" – Third Edition- Packt publications

SRI A.S.N.M. GOVERNMENT COLLEGE (A), PALAKOL

CBCS/Semester System
(W.e.f. 2020-21 Admitted Batch)

B.Sc (Data Science)
II YEAR IV SEMESTER

DATA VISUALIZATION LAB USING TABLEAU

1. Connect to data Sources
2. Create Univariate Charts
3. Create Bivariate and Multivariate charts
4. Create Maps
5. Calculate user-defined fields
6. Create a workbook data extract
7. Save a workbook on a Tableau server and web
8. Export images, data.

SRI A.S.N.M. GOVERNMENT COLLEGE (A), PALAKOL

CBCS/Semester System
(W.e.f. 2020-21 Admitted Batch)

B.Sc (Data Science)
II YEAR IV SEMESTER

DATA VISIUALISATION
Model Question Paper

Max. Marks: 75 **Time: 3Hrs Max**

SECTION-A **5x5=25M**
(Answer any Five of the following)

1. Explain creating visual analytics with tableau desktop.
2. Discuss bar chart, line chart, area fill and pie chart with examples.
3. What are calculated values and table calculations?
4. Explain how you plot your own locations on a map.
5. How views are edited in tableau server 6.What are generated values? Discuss
7. What is the usage of Gantt charts? Explain with examples
8. Discuss table calculation functions

SECTION-B

5 X10=50M

9. Explain how to blend different data sources in a single work sheet

(OR)

Discuss how different tables are joined with tableau.

10. Discuss how to work with filters to enhance views

(OR)

What are different set groups and hierarchies in visualization.

11. What is aggregation explain how dialogue box is created using calculations.

(OR)

Discuss how formulas are build using table calculations

12. Discuss how to create a standard map view with an example

(OR)

Explain how data shaping is done to enable point to point mapping 13.How self evidence ad-hoc analyses is provided with parameters.

(OR)

Explain methods or generating new data with fore caste