

Department of Statistics

Project in academic year 2023-24

Sr. No.	Title of the project	No. of Students	Project Guide
1.	Analysis of the Prevalence and Health Impact of PCOS or PCOD Among Women	4	Mr. Bhosale A. B.
2.	Bank Customer Churn Tendency in Foreign Countries Using Machine Learning	3	Mr. Pawar A. A.
3.	Discovering a Hotel Selection Preference of Youth: A Statistical Approach	2	Mr. Pawar A. A.
4.	Analysis of Indian Crop Production Data Using Advanced Statistical Tools	3	Mr. Tangawade A. S
5.	Skill-Based Job Recommendation System Using Machine Learning Techniques	3	Ms. Pandhare R. S.
6.	Statistical Analysis of Healthcare Insurance	2	Ms. Patil R. M.
7.	Statistical Analysis of Caesarean Section for Pregnant Women	3	Ms. Bhosale A. K.



“Dissemination of Education for Knowledge, Science and Culture”
-Shikshanmaharshi Dr. Bapuji Salunkhe



(स्वायत्त) कोल्हापूर

**VIVEKANAND COLLEGE
KOLHAPUR
(Empowered Autonomous)**

DEPARTMENT OF STATISTICS

A PROJECT REPORT

on

**“Analysis of the Prevalence and
Health Impact of PCOS or PCOD
Among Women”**

Submitted by

Miss. Redekar Divyarani Pandurang

Miss. Kumbhar Shamal Sambhaji

Miss. Patil Priyanka Sayaji

Mrs. Chavan Shubham Suresh

in partial fulfillment for the award of the degree of

MASTER OF SCIENCE

in

STATISTICS

2023-24

CERTIFICATE

This is to Certify that,


Sr. No.	Name	Roll No.
1	Ms. Redekar Divyarani Pandurang	1421
2	Ms. Kumbhar Shamal Sambhaji	1411
3	Ms. Patil Priyanka Sayaji	1419
4	Mr. Chavan Shubham Suresh	1402

Have satisfactorily completed the project work on “**Analysis of the Prevalence and Health Impact on PCOS or PCOD Among Women**” as a part of practical evaluation course for **M.Sc. II**, prescribed by the Department of Statistics, *Vivekanand College, Kolhapur (Empowered Autonomous)* in the academic year **2023-24**.

This project has been completed under our guidance and supervision. To the best of our knowledge and belief, the matter presented in this project report is original and has not been submitted elsewhere for any other purpose.

Date:

Place: Kolhapur


Project Guide

(Mr. Bhosale. A. B.)


Examiner


Head

(Mrs. Shinde.V.C.)

HEAD
DEPARTMENT OF STATISTICS
VIVEKANAND COLLEGE, KOLHAPUR
(EMPOWERED AUTONOMOUS)

OBJECTIVES

- ✓ To study the prevalence of PCOD/PCOS in women on irregular period.
- ✓ To predict the women's with PCOD/PCOS based risk factor.
- ✓ To determine which reasons are most responsible for PCOD/PCOS.
- ✓ To check the classification of PCOD/PCOS in women's using various machine learning algorithms.

MAJOR FINDING

- Maximum number of women experienced problems in pregnancy due to PCOD/PCOS.
- PCOD appears to be more prevalent among younger age groups, with the highest number of cases observed in the age group 20-25.
- Approximately 90% of women having PCOD/PCOS consumed fast food, oily/spicy food.
- Age affects the occurrence of PCOD or PCOS, while it does not affect the types of periods.
- Marital status affects occurrence of PCOD/PCOS.
- Every 22 of 100 women face problem of PCOD/PCOS.
- Random forest model gives highest (86%) accuracy of classification, whereas logistic regression gives 81% accuracy, the AUC is 0.91 which indicates that discrimination is fair.
- From principal component analysis, 12 number of components from 27 component are extracted which covers maximum variation within dataset.

Limitations of the Study

- Tools applied to the data can change their performance if we change the data
- We develop model only with available variables but if add other important variables then we expect that our models give better results

“Dissemination of Education for Knowledge, Science and Culture”

-Shikhanmaharshi Dr. Bapuji Salunkhe



VIVEKANAND COLLEGE, KOLHAPUR
(Empowered Autonomous)

DEPARTMENT OF STATISTICS

A PROJECT REPORT

On

BANK CUSTOMER CHURN TENDENCY IN FOREIGN COUNTRIES
USING MACHINE LEARNING

Submitted by

Ms. Mane Supriya Narayan

Ms. Patil Asavari Ajit

Mr. Mulani Sabir Javed

in partial fulfillment for the award of the

degree of

MASTER OF SCIENCE

in

STATISTICS

2023-24

CERTIFICATE

This is to Certify that,

Sr. No.	Name	Roll No.
1	Ms. Mane Supriya Narayan	1412
2	Ms. Patil Asavari Ajit	1415
3	Mr. Mulani Sabir Javed	1413

Have satisfactorily completed the project work on “**BANK CUSTOMER CHURN TENDENCY IN FOREIGN COUNTRIES USING MACHINE LEARNING**” as a part of Practical Evaluation for **M.Sc. II**, prescribed by the Department of Statistics, *Vivekanand College, Kolhapur (Empowered Autonomous)* in the academic year **2023-24**.

This project has been completed under our guidance and supervision. To the best of our knowledge and belief, the matter presented in this project report is original and has not been submitted elsewhere for any other purpose.



Project Guide

(Mr. Pawar. A. A)



Examiner

Head

(Ms. Pawar V. V.)

OBJECTIVES

- To identify & visualize which factors contribute to customer churn.
- To build a prediction model that will classify if a customer is going to churn or not.
- To compare the performance of the fitted churn models & find the best model for prediction of customer churn.

CONCLUSION

- The variables such as CreditScore, EstimatedSalary are independent of customer churning.
- The features such as Age of customer, Balance, Tenure, NumOfProducts are most important features responsible for customer churning.
- On the basis of the Accuracy given by each of the models Random Forest has highest accuracy of about 82.45% for the prediction of customer churning.
- Comparative study of the models resulted that the Random Forest has the highest Accuracy of 82.45%, also the greatest Area Under Curve about 92%. So, Random Forest is the best model for customer churn prediction.

**“Dissemination of Education for Knowledge, Science and Culture”
-Shikshanmaharshi Dr. Bapuji Salunkhe**



(स्वायत्त) कोल्हापूर

**VIVEKANAND COLLEGE
KOLHAPUR
(Empowered Autonomous)**

DEPARTMENT OF STATISTICS

A PROJECT REPORT

on

**“Discovering a Hotel Selection Preference
of Youth: A Statistical Approach”**

Submitted by

**Miss. Gurav Sanjivani Sadanand
Miss. Chougale Swati Pandit**

*in partial fulfillment for the award of
the degree of*

MASTER OF SCIENCE

in

**STATISTICS
2023-24**

CERTIFICATE

This is to Certify that,


Sr. No.	Name	Roll No.
1	Miss. Gurav Sanjivani Sadanand	1407
2	Miss. Chougale Swati Pandit	1403


Have satisfactorily completed the project work on “**Discovering a Hotel Selection Preference of Youth: A Statistical Approach**” as a part of practical evaluation course for **M.Sc. II**, prescribed by the Department of Statistics, *Vivekanand College, Kolhapur (Empowered Autonomous)* in the academic year **2023-24**.


This project has been completed under our guidance and supervision. To the best of our knowledge and belief, the matter presented in this project report is original and has not been submitted elsewhere for any other purpose.

Date: 29/05/2024

Place: Kolhapur


Project Guide
(Mr. Pawar A.A.)


20/05/24
Examiner


Head
(Mrs. Shinde V.C.)
HEAD
DEPARTMENT OF STATISTICS
VIVEKANAND COLLEGE, KOLHAPUR
(EMPOWERED AUTONOMOUS)

OBJECTIVES

- ✓ To explore the customer preferences.
- ✓ To explore the preference of booking platforms.
- ✓ To understand the importance of Amenities and Factors preference.

OVERALL CONCLUSIONS

- Customers prefer parking, complimentary breakfast, room service and free wi-fi when choosing a hotel. They also value food taste, customer service and location highly.
- Essential amenities include parking, complimentary breakfast, gym/pool, pet-friendly options, laundry service, and spa.
- Most of the customers prefer Early Booking Discounts strategy for booking the hotel.
- There is an association between Food type and Gender.
- Level of standards in Amenities as well as factors is most important.

“Dissemination of Education for Knowledge, Science and Culture”
-Shikhanmaharshi Dr. Bapuji Salunkhe



VIVEKANAND COLLEGE
KOLHAPUR

(Empowered Autonomous)

DEPARTMENT OF STATISTICS

**A PROJECT REPORT
ON**

**“Analysis of Indian Crop Production Data Using Advanced
Statistical Tools”**

Submitted by

**Mr. Chougale Yuvaraj Jeevan
Ms. Dhere Asmita Nandkumar
Ms. Kalekar Manasi Manohar**

*in partial fulfillment for the award of
the degree of*

MASTER OF SCIENCE

in

STATISTICS

2023-24

CERTIFICATE

This is to Certify that,

Sr. No.	Name	Roll No.
1	Mr. Chougale Yuvaraj Jeevan.	1404
2	Ms. Dhere Asmita Nandkumar.	1405
3	Ms. Kalekar Manasi Manohar.	1408

Have satisfactorily completed the project work on “**Analysis of Indian Crop Production Data using advanced statistical tools**” as a part of practical evaluationenhancement course for **M.Sc. II**, prescribed by the Department of Statistics, *Vivekanand College, Kolhapur (Empowered Autonomous)* in the academic year **2023-24**.

This project has been completed under our guidance and supervision. To the best of our knowledge and belief, the matter presented in this project report is original and has not been submitted elsewhere for any other purpose.

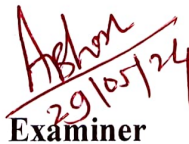
Date

Place: Kolhapur



Project Guide

(Mr. Tangavade A. S.)



Examiner



Head

(Mrs. Shinde V. C.)

HEAD
DEPARTMENT OF STATISTICS
VIVEKANAND COLLEGE, KOLHAPUR
(EMPOWERED AUTONOMOUS)

OBJECTIVES

- To find most cultivated crops in India.
- Exploring correlations between crop production and various factors such as climate variables, soil characteristics and agricultural practices.
- To study the impact of the weather condition on the crop yield.
- To build the model for the prediction of crop yield from various factors.
- Comparative study of performance of multiple linear regression (MLR) and random forest regression (RFR)



OVERALL CONCLUSIONS

- The descriptive statistics indicate that rice and wheat are the predominant crops cultivated in India. Moreover, rice, cotton, and soybean are primarily cultivated during the kharif season, while wheat is predominantly grown during the rabi season. This highlights the seasonal cultivation patterns observed in India, with specific crops aligned to the distinct climatic conditions of each season.
- The correlation analysis reveals notable associations between crop yield and key factors such as nitrogen level, pH level, rainfall, and temperature. It underscores the significant influence that these variables exert on crop productivity. Specifically, higher nitrogen levels, optimal pH conditions, adequate rainfall, and suitable temperatures are found to positively impact crop yield, emphasizing the importance of considering these factors in agricultural management practices.
- From performance measures like Mean Squared Error (MSE), Root Mean Squared Error (RMSE) and coefficient of determination (R^2), we can conclude that performance of RFR model is better than MLR model. This result helping to inform model selection and decision-making in agricultural forecasting and planning.

“Dissemination of Education for Knowledge, Science and Culture.”

-Shikshanmaharshi Dr. Bapuji Salunkhe



(स्वायत्त) कोल्हापूर

VIVEKANAND COLLEGE,

KOLHAPUR

(Empowered Autonomous)

DEPARTMENT OF STATISTICS

A PROJECT REPORT

ON

**“Skill-Based Job Recommendation System Using
Machine Learning Techniques.”**

Submitted by

Mr. Chavan Shubham Suresh

Mr. Parit Amol Dhanaji

Ms. Alavane Rutuja Maruti

*In partial fulfillment for the award
of the degree of*

MASTER OF SCIENCE

in

STATISTICS

2023-24

“Dissemination of Education for Knowledge, Science and Culture.”

-Shikshanmaharshi Dr. Bapuji Salunkhe



(स्वायत्त) कोल्हापूर

VIVEKANAND COLLEGE, KOLHAPUR

(Empowered Autonomous)

DEPARTMENT OF STATISTICS

CERTIFICATE

This is to Certify that,

Sr. No.	Name	Roll No.
1	Mr. Chavan Shubham Suresh	1402
2	Mr. Parit Amol Dhanaji	1414
3	Ms. Alavane Rutuja Maruti	1401

Have satisfactorily completed the project work on “*Skill-Based Job Recommendation System using machine learning techniques.*” as a part of practical evaluation course for *M.Sc. II*, prescribed by the Department of Statistics, *VivekanandCollege, Kolhapur (Empowered Autonomous)* in the academic year **2023-24**.

This project has been completed under our guidance and supervision. To the best of our knowledge and belief, the matter presented in this project report is original and has not been submitted elsewhere for any other purpose.


29/05/2024
Project Guide

(Ms. Pandhare R. S.)


29/05/24
Examiner

Head

(Mrs. Shinde V. C.)

OBJECTIVES

- To build a recommendation system that will help candidates to get suitable jobs concerning their skills.
- To suggest skills to the candidates so that they can acquire them and get a suitable job.
- To check which factors are significant and which factors affect our target variable.
- To check which type of Job candidates can prefer the most among the job type viz. Data Scientist, Data Analyst, and Data Engine

CONCLUSION

- ❖ From our analysis , we can suggest that the most important skills for the candidates are SQL, Python, R, and Machine learning so that they can acquire them and get a suitable job.
- ❖ The Random Forest model is the best fit model for our data.
- ❖ We checked the significant factors are Salary, No. of_Skills, Reviews, Rating, Date_Since_Posted, Revenue, Employees and location doesn't affect our target variable.
- ❖ Most of the candidates prefer python skill.
- ❖ Salary is dependent on Number of skills.
- ❖ Data Scientist is the most prefer job by candidates.
- ❖ For Data Scientist, the most preferable skills are SQL, python, and tableau. For Data Analyst, the key skills are python, SQL, and Hadoop. For Data Engineer, key skills are python SQL and Machine Learning.

“Dissemination Education for Knowledge, Science and Culture”

-Shikhanmaharshi Dr. Bapuji Salunkhe



(स्वायत्त) कोल्हापूर

VIVEKANAND COLLEGE, KOLHAPUR

(Empowered Autonomous)

DEPARTMENT OF STATISTICS

A PROJECT REPORT

On

**“STATISTICAL ANALYSIS ON HEALTHCARE
INSURANCE”**

Submitted by

Ms. Patil Priti Babasaheb

Ms. Ekal Vedika Shivanand

In partial fulfillment for the award of

the degree of

MASTER OF SCIENCE

In

STATISTICS

2023-24

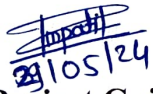
CERTIFICATE

This is to Certify that,

Sr. No.	Name	Roll No.
1	Ms. Patil Priti Babasaheb	1418
2	Ms. Ekal Vedika Shivanand	1406

Have satisfactorily completed the project work on “**Statistical Analysis On Healthcare Insurance**” as a part of practical evaluation for **M.Sc. II**, prescribed by the Department of Statistics, *Vivekanand College, Kolhapur (Empowered Autonomous)* in the academic year **2023-24**.

This project has been completed under our guidance and supervision. To the best of our knowledge and belief, the matter presented in this project report is original and has not been submitted elsewhere for any other purpose.


29/05/24

Project Guide

(Ms. Patil R.M.)


29/05/24

Examiner

Head

(Ms. Pawar V. V.)

OBJECTIVES

- To identify key factors affecting on medical charges.
- To investigate the relationship between BMI and medical charges.
- To examine the impact of smoking on medical expenses.
- To analyze demographic influences (age, sex, and number of children) on charges.
- To assess regional variations in medical charges.

CONCLUSION

The analysis of the provided dataset reveals several insights into the factors affecting medical charges

- ✓ Age & BMI Both have strong positive correlations with medical charges, meaning older individuals and those with higher BMI tend to have higher medical expenses.
- ✓ Smoking is a major factor that significantly increases medical charges.
- ✓ While age and BMI are strong influencers, the impact of sex.
- ✓ The interplay of these factors (age, BMI, smoker status, region) creates a complex picture of medical charges, underscoring the need for a comprehensive approach in understanding and predicting medical expenses.

“Education for Knowledge, Science and Culture”

-Shikshanmaharshi Dr. Bapuji Salunkhe



(स्वायत्त) कोल्हापूर

VIVEKANAND COLLEGE KOLHAPUR

(Empowered Autonomous)

DEPARTMENT OF STATISTICS

A PROJECT REPORT

on

“STATISTICAL ANALYSIS OF CAESAREAN SECTION FOR
PREGANANT WOMEN”

Submitted by

Ms. Pradnya Bharat Kamble

Ms. Lalita Ramdas Koli

Mr. Ketan Kiran Patil

*in partial fulfillment for the award of
the degree of*

MASTER OF SCIENCE

in

STATISTICS

2023-24

CERTIFICATE

This is to Certify that,

Sr. No.	Name	Roll No.
1	Ms. Kamble Pradnya Bharat	1409
2	Ms. Koli Lalita Ramdas	1410
3	Mr. Patil Ketan Kiran	1417

Have satisfactorily completed the project work on “**Statistical Analysis of Caesarean Section For Pregnant Women**” as a part of practical evaluation course for **M.Sc. II**, prescribed by the Department of Statistics, *Vivekanand college, Kolhapur (Empowered autonomous)* in the academic year **2023-24**.

This project has been completed under our guidance and supervision. To the best of our knowledge and belief, the matter presented in this project report is original and has not been submitted elsewhere for any other purpose.

Date:

Place: Kolhapur

*for
Bhosale*

Project Guide

(Ms. Bhosale. A. K.)

*Asker
29/05/24*
Examiner

W. Pawar

Head

(Ms. Pawar V. V.)

HEAD
DEPARTMENT OF STATISTICS
VIVEKANAND COLLEGE, KOLHAPUR
(EMPOWERED AUTONOMOUS)

Objectives

- 1) To see caesarean delivery is suitable according to antenatal & postnatal condition.
- 2) To determine which reasons are most responsible for caesarean Section delivery.
- 3) To identify the trends and factors associated with caesarean section.
- 4) To check the classification of caesarean section in women's using various machine learning algorithm.

Major findings

- ❖ Maximum number of Women's experienced problems in delivery due to Age .
- ❖ Approximately 85% of women need caesarean delivery due to advancing Age.
- ❖ Decision tree model gives highest (83%) accuracy of classification, whereas random forest model gives 82% accuracy, the AUC is 0.93 which indicates that discriminant is fair.