

Department of Artificial Intelligence and Data Science

**ZEAL COLLEGE OF ENGINEERING & RESEARCH,
NARHE
Department of Artificial Intelligence and Data Science**



DEPARTMENT OF Artificial Intelligence and Data Science

317535: Internship

ACADEMIC YEAR 2022-23

“Data Science And Business Analytics”

NAME: JADHAV SHUBHAM

ROLL NO: T1411026

Department Of Artificial Intelligence and Data Science



ZEAL COLLEGE OF ENGINEERING & RESEARCH, NARHE

CERTIFICATE

This is to certify that the Internship Entitled

“Data Science and Business Analytics”

Submitted by

Student Name: JADHAV SHUBHAM

Exam No: T1411026

Mr. Dikshendra Sarpate
H. O. D
Dept. of AIDS

Dr. Ajit Kate
Principal
ZCOER

Signature of Internal Internship Supervisor

Signature of External Internship Supervisor

Internship Certificate



Internship Place Details

Company background-Organization	OASIS INFOBYTE
Activities/Scope	Only for School and College ERP (Enterprise Resource Planning) Software and Technical Workshops for Entrepreneurship Development
Objective of Study	<ul style="list-style-type: none">• Learning Data Science• Improving Python programming skills• Gaining experience with database management• Understanding software development life cycle• Building a portfolio• Gaining Industry level Knowledge
Supervisor Details (Name, Designation, Company Name, Email_id, Contact number)	<ul style="list-style-type: none">• Girish Pawar• Founder and CEO• official@oasisinfobyte.com• 7030020973

Index

Sr. No	Chapter Name	Page No.
1	Introduction	6
2	Title/Problem statement/Objective	7
3	Motivation/Scope and rationale of the study	8
4	Weekly Activity Report	9
5	Methodological details	16
5	Results/Analysis/Interferences and Conclusion	17
7	Suggestions	18
8	Acknowledgement	19
9	List of References	20
10	Evaluation Sheet	21

Introduction

In today's rapidly changing retail industry, it's becoming increasingly important for stores to personalize the shopping experience for their customers. One way to achieve this is by offering personalized product recommendations based on a customer's preferences, buying history, and behavior.

Artificial Intelligence & Machine Learning Offerings Help Organizations Build Highly-Customized Solutions Running On Advanced Machine Learning Algorithms. Our Machine Learning Services Offer Advanced Algorithms To Help Organizations In Solving Key Business Challenges, Enabling Data-Driven Decision Making And Creating Innovative Business Models.

Remote Data Science Internships Are A Unique Chance To Gain Experience In The Midst Of The Virtual Workforce While Remaining Immersed In One Of The Top Organizations In The Field. Data Analysis Internships Are Some Of The Most Competitive And Popular Within The Broader Data Science Field.

Title/Problem Statement/Objective

Title:-

Data Science And Business Analytics

Problem Statement:-

To Build a model for unemployment in India.

Objective:-

1. Data Acquisition: Collecting the Iris dataset and obtaining relevant unemployment data for India from credible sources such as government publications or official websites.
2. Data Preprocessing: Cleaning and preparing the acquired data for analysis, including handling missing values, removing duplicates, and transforming the data into a suitable format for further processing.
3. Exploratory Data Analysis: Conducting a comprehensive analysis of the unemployment dataset, exploring various factors such as unemployment rates, demographic variables, education levels, and regional variations. Identifying key trends and patterns in the data.
4. Visualization: Creating informative and visually appealing charts, graphs, and interactive visualizations using Python libraries such as Matplotlib, Seaborn, or Plotly. Presenting the unemployment trends across different time periods, states, and demographic groups to effectively communicate the findings.
5. Statistical Analysis: Applying statistical techniques to gain deeper insights into the data, such as calculating the average unemployment rate, identifying outliers, examining correlations between variables, and conducting hypothesis testing if required.
6. Predictive Modeling (optional): Building predictive models, such as time series forecasting or regression analysis, to predict future unemployment rates based on historical data. This step can provide additional insights into future trends and assist policymakers in making informed decisions.
7. Documentation and Reporting: Documenting the entire project, including the data sources, preprocessing steps, analysis techniques, and visualizations. Presenting the findings in a clear and concise manner through a final project report or interactive dashboard, ensuring that the results are easily understandable and accessible to a wide audience.

Motivation/Scope & Rationale of the Study

Motivation :-

Title: Analyzing Iris Unemployment in India: Empowering Change through Python

Introduction:

In today's rapidly evolving world, data plays a crucial role in decision-making and driving positive change. Unemployment remains a significant challenge faced by societies worldwide, and India is no exception. To address this issue, we propose a Python project focused on analyzing and visualizing unemployment data in India using the renowned Iris dataset. Through this project, we aim to empower individuals, policymakers, and organizations with valuable insights into the unemployment situation, enabling them to make informed decisions and take meaningful action to tackle this pressing issue.

1. Understanding the magnitude of the problem:

The first step towards resolving any problem is to gain a comprehensive understanding of its magnitude. By working on this Python project, you will have the opportunity to explore the Iris dataset, which provides a rich collection of unemployment data in India. Analyzing this dataset will allow you to grasp the severity of the unemployment crisis, identify trends, and gain insights into the factors contributing to unemployment rates in different regions and demographics.

2. Unveiling the underlying causes:

Unemployment is a complex issue with various underlying causes, such as economic factors, education, skill gaps, industry-specific challenges, and social dynamics. By leveraging Python's data analysis capabilities, you can dive deep into the Iris dataset and uncover correlations, patterns, and dependencies. This exploration will enable you to identify the primary causes of unemployment in India, helping stakeholders understand the root problems and devise effective strategies to address them.

3. Visualizing the data for impactful insights:

Data visualization is a powerful tool for conveying complex information in an intuitive and impactful manner. Through this project, you will harness Python's visualization libraries, such as Matplotlib and Seaborn, to create stunning visual representations of unemployment data in India. By presenting the data in visually appealing and easily digestible formats, you can communicate your findings effectively, increasing awareness and generating support for initiatives aimed at reducing unemployment.

4. Predictive modeling for proactive interventions:

Python offers an array of machine learning libraries, including scikit-learn and TensorFlow, which can

be leveraged to build predictive models based on historical unemployment data. By training these models with the Iris dataset, you can forecast future unemployment trends, identify potential hotspots, and develop proactive intervention strategies. Armed with this information, policymakers and organizations can implement targeted programs, training initiatives, and employment opportunities to address unemployment issues before they escalate.

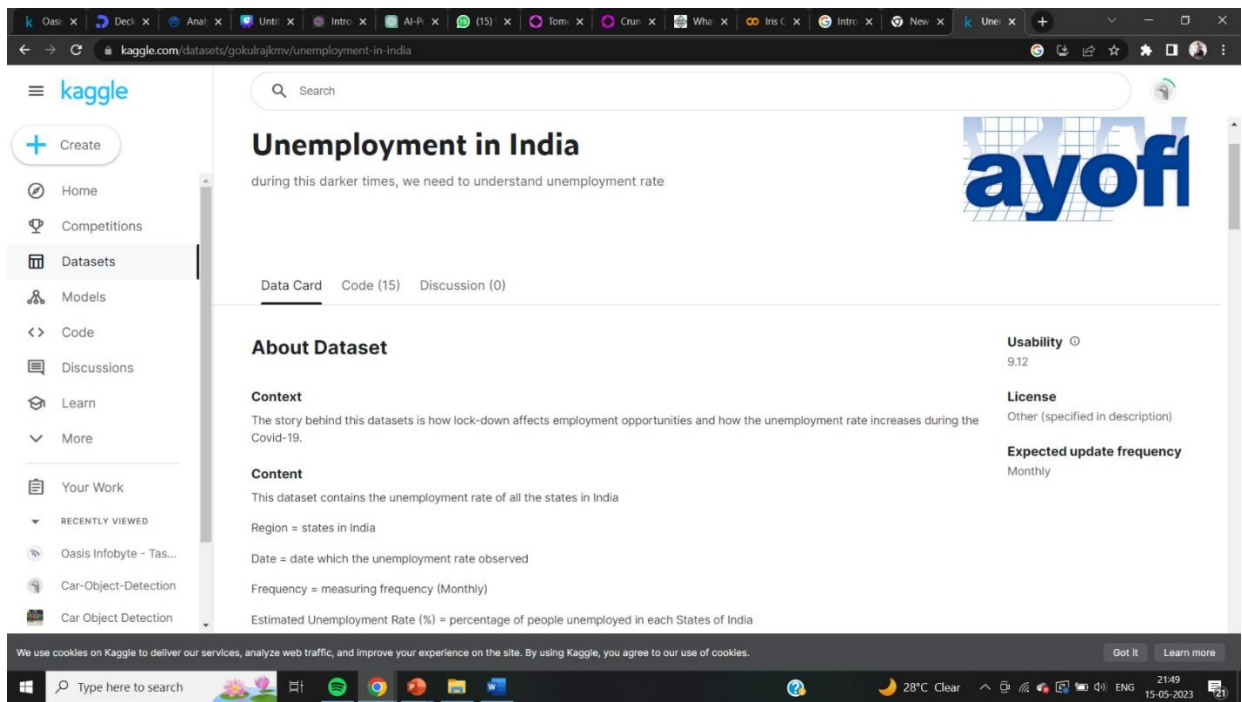
5. Driving change through data-driven decision-making:

The ultimate goal of this Python project is to drive change by equipping stakeholders with actionable insights and evidence-based decision-making tools. Your work on analyzing the Iris unemployment dataset will contribute to empowering policymakers, governmental organizations, non-profits, and social activists to implement effective policies and interventions that lead to tangible improvements in the employment scenario in India. By utilizing Python's capabilities, you can be part of a movement that aims to bring about positive social and economic transformation.

Weekly Activity Report

- **Week 1:-**

DATA COLLECTION :



- **Week 2 :**

DATA PREPARATION :

Steps:-

Data cleaning:

Raw data may contain errors, outliers, or missing values, which can negatively impact the accuracy of the analysis. Thus, it is important to clean the data by removing or correcting errors, handling outliers, and imputing missing values.

Feature engineering:

Feature engineering involves creating new variables that are more informative for the analysis. In this case, the number of study hours may be the only feature available. However, additional features such as the student's demographic information or past academic performance may be useful in predicting academic performance.

Data splitting:

The data needs to be split into training and testing sets. The training set is used to build the predictive

model, while the testing set is used to evaluate the model's performance.

Data normalization:

Data normalization is the process of scaling the data to a common range. Normalization ensures that variables with larger values do not dominate the analysis. This step is particularly important when using machine learning models such as linear regression or neural networks.

Region	Date	Frequency	Estimated	Estimated Region	longitude	latitude
Andhra Pradesh	31-01-202	M	5.48	16635535	41.02 South	15.9129
Andhra Pradesh	29-02-202	M	5.83	16545652	40.9 South	15.9129
Andhra Pradesh	31-03-202	M	5.79	15881197	39.18 South	15.9129
Andhra Pradesh	30-04-202	M	20.51	11336911	33.1 South	15.9129
Andhra Pradesh	31-05-202	M	17.43	12988845	36.46 South	15.9129
Andhra Pradesh	30-06-202	M	3.31	19805400	47.41 South	15.9129
Andhra Pradesh	31-07-202	M	8.34	15431615	38.91 South	15.9129
Andhra Pradesh	31-08-202	M	6.96	15251776	37.83 South	15.9129
Andhra Pradesh	30-09-202	M	6.4	15220312	37.47 South	15.9129
Andhra Pradesh	31-10-202	M	6.59	15157557	37.34 South	15.9129
Assam	31-01-202	M	4.66	13051904	52.98 Northeast	26.2006
Assam	29-02-202	M	4.41	10088268	40.77 Northeast	26.2006
Assam	31-03-202	M	4.77	11542888	46.73 Northeast	26.2006
Assam	30-04-202	M	11.06	6830817	29.55 Northeast	26.2006
Assam	31-05-202	M	9.55	11367897	48.26 Northeast	26.2006
Assam	30-06-202	M	0.6	9095944	35.07 Northeast	26.2006
Assam	31-07-202	M	3.77	10286757	40.88 Northeast	26.2006
Assam	31-08-202	M	5.53	9781310	39.52 Northeast	26.2006
Assam	30-09-202	M	1.19	14107641	54.38 Northeast	26.2006
Assam	31-10-202	M	3.02	11949329	46.84 Northeast	26.2006
Bihar	31-01-202	M	10.61	26397671	37.72 East	25.0961
Bihar	29-02-202	M	10.29	26281655	37.35 East	25.0961
Bihar	31-03-202	M	15.43	25717519	38.69 East	25.0961
Bihar	30-04-202	M	46.64	16046236	38.17 East	25.0961
Bihar	31-05-202	M	45.96	16280203	38.16 East	25.0961
Bihar	30-06-202	M	17.82	23963933	36.86 East	25.0961
Bihar	31-07-202	M	17.79	24989370	36.14 East	25.0961

- **Week 3 & 4 :**
- **Model Planning & Model Building**

Model planning involves choosing a suitable regression model, selecting relevant features, tuning hyperparameters, training the model, evaluating its performance using metrics such as MSE or R-squared, and fine-tuning the model as needed.

Department of Artificial Intelligence and Data Science

The image displays two screenshots of a Google Colab notebook titled "Unemployment in India.ipynb".

Top Screenshot: Shows the initial code and the first view of the data. The code imports necessary libraries and reads a CSV file. The data is displayed as a table with the following columns: Region, Date, Frequency, Estimated Unemployment Rate (%), Estimated Employed, Estimated Labour Participation Rate (%), Region.1, longitude, and latitude. The first few rows show data for Andhra Pradesh, and the last row shown is for West Bengal.

	Region	Date	Frequency	Estimated Unemployment Rate (%)	Estimated Employed	Estimated Labour Participation Rate (%)	Region.1	longitude	latitude
0	Andhra Pradesh	31-01-2020	M	5.48	16635535	41.02	South	15.9129	79.740
1	Andhra Pradesh	29-02-2020	M	5.83	16545652	40.90	South	15.9129	79.740
2	Andhra Pradesh	31-03-2020	M	5.79	15881197	39.18	South	15.9129	79.740
3	Andhra Pradesh	30-04-2020	M	20.51	11336911	33.10	South	15.9129	79.740
4	Andhra Pradesh	31-05-2020	M	17.43	12988845	36.46	South	15.9129	79.740
...
262	West Bengal	30-06-2020	M	7.29	30726310	40.39	East	22.9868	87.855
263	West Bengal	31-07-2020	M	6.83	35372506	46.17	East	22.9868	87.855
264	West Bengal	31-08-2020	M	14.87	33298644	47.48	East	22.9868	87.855
265	West Bengal	30-09-2020	M	9.35	35707239	47.73	East	22.9868	87.855

Bottom Screenshot: Shows the same notebook after running `df.describe()`. It displays summary statistics for the data. The statistics are as follows:

	Estimated Unemployment Rate (%)	Estimated Employed	Estimated Labour Participation Rate (%)	longitude	latitude
count	267.000000	2.670000e+02	267.000000	267.000000	267.000000
mean	12.236929	1.396211e+07	41.681573	22.826048	80.532425
std	10.803283	1.336632e+07	7.845419	6.270731	5.831738
min	0.500000	1.175420e+05	16.770000	10.850500	71.192400
25%	4.845000	2.838930e+06	37.265000	18.112400	76.085600
50%	9.650000	9.732417e+06	40.390000	23.610200	79.019300
75%	16.755000	2.187869e+07	44.055000	27.278400	85.279900
max	75.850000	5.943376e+07	69.690000	33.778200	92.937600

Below the statistics, the output of `df.isnull().sum()` is shown, indicating that there are no missing values in any of the columns.

Region 0
Date 0
Frequency 0
Estimated Unemployment Rate (%) 0
Estimated Employed 0
Estimated Labour Participation Rate (%) 0
Region.1 0
longitude 0
latitude 0
dtype: int64

A OneDrive notification is visible in the bottom right corner: "Screenshot saved. The screenshot was added to your OneDrive."

Department of Artificial Intelligence and Data Science

The image displays two screenshots of a Jupyter Notebook titled "Unemployment in India.ipynb" running in a web browser. The notebook is open to the "Code" tab, showing Python code for data analysis.

First Screenshot:

```
df = df.rename(columns=[df.columns[0]:'State',df.columns[3]:'EUR',df.columns[4]:'EE',df.columns[5]:'ELPR',df.columns[6]:'Region'])
df.head()
```

	State	Date	Frequency	EUR	EE	ELPR	Region	longitude	latitude
0	Andhra Pradesh	31-01-2020	M	5.48	16635535	41.02	South	15.9129	79.74
1	Andhra Pradesh	29-02-2020	M	5.83	16545652	40.90	South	15.9129	79.74
2	Andhra Pradesh	31-03-2020	M	5.79	15881197	39.18	South	15.9129	79.74
3	Andhra Pradesh	30-04-2020	M	20.51	11336911	33.10	South	15.9129	79.74
4	Andhra Pradesh	31-05-2020	M	17.43	12988845	36.46	South	15.9129	79.74

```
[ ] df["State"].unique()
array(['Andhra Pradesh', 'Assam', 'Bihar', 'Chhattisgarh', 'Delhi', 'Goa',
      'Gujarat', 'Haryana', 'Himachal Pradesh', 'Jammu & Kashmir',
      'Jharkhand', 'Karnataka', 'Kerala', 'Madhya Pradesh',
      'Maharashtra', 'Meghalaya', 'Odisha', 'Puducherry', 'Punjab',
      'Rajasthan', 'Sikkim', 'Tamil Nadu', 'Telangana', 'Tripura',
      'Uttar Pradesh', 'Uttarakhand', 'West Bengal'], dtype=object)

[ ] df["Region"].unique()
array(['South', 'Northeast', 'East', 'West', 'North'], dtype=object)

[ ] df.groupby("Region").size()
Region
East      40
```

Second Screenshot:

```
[ ] df.groupby("Region").size()
Region
East      40
North     79
Northeast 38
South     60
West      50
dtype: int64

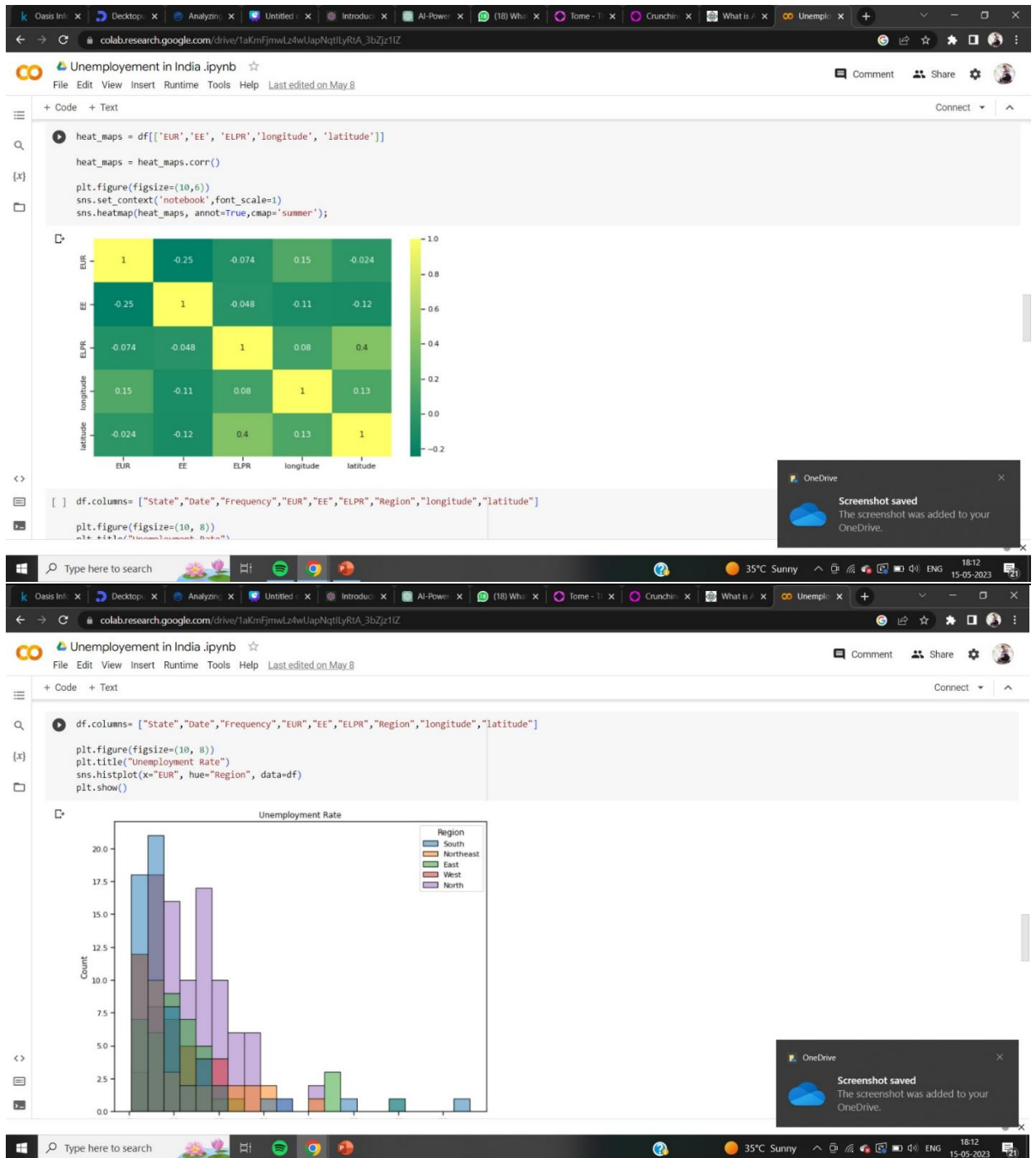
region_stats = df.groupby(['Region'])[['EUR','EE','ELPR']].mean().reset_index()
region_stats = round(region_stats,2)
region_stats
```

	Region	EUR	EE	ELPR
0	East	13.92	19602366.90	40.11
1	North	15.89	13072487.92	38.70
2	Northeast	10.95	3617105.53	52.06
3	South	10.45	14040589.33	40.44
4	West	8.24	18623512.72	41.26

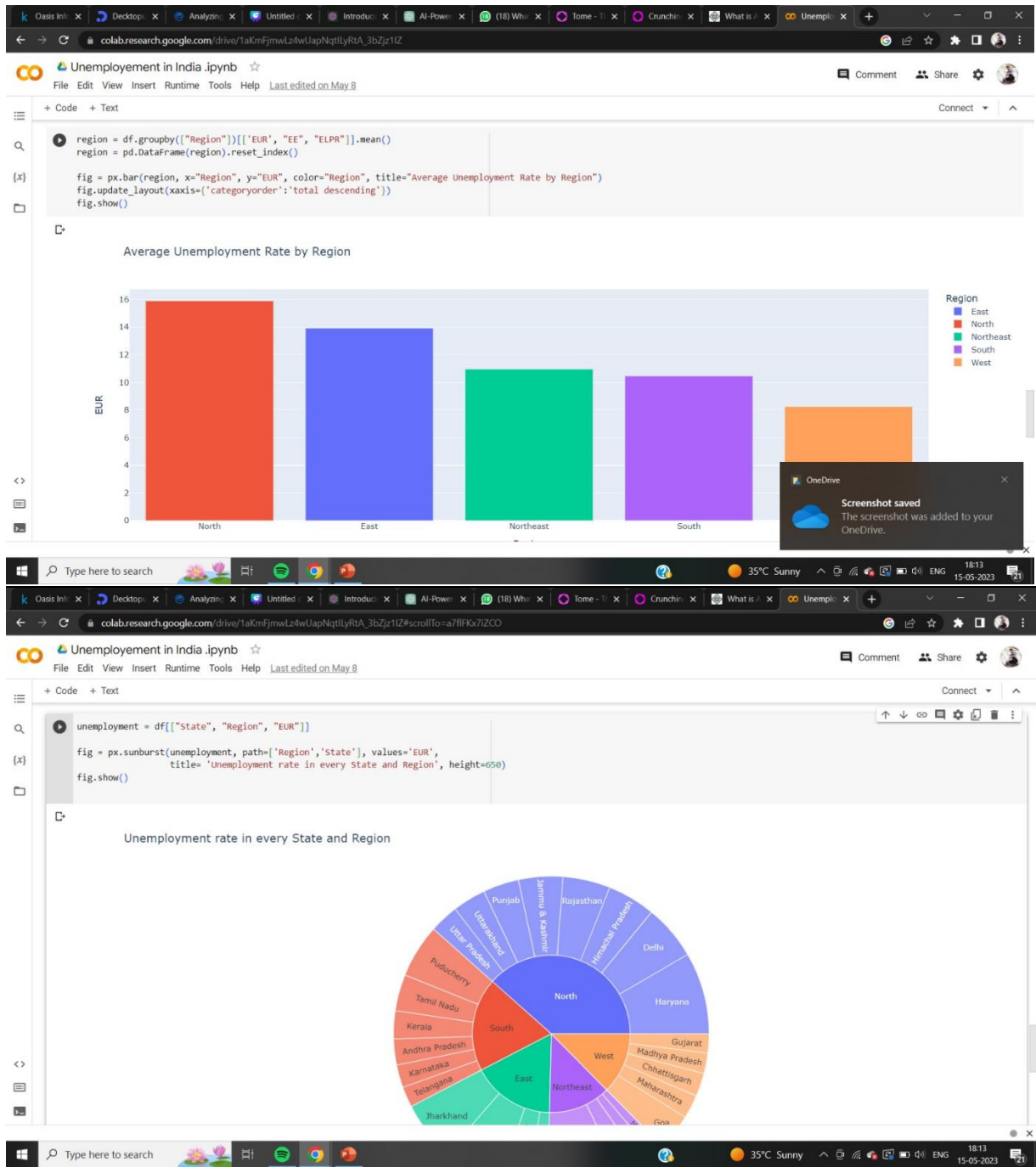
```
[ ] heat_maps = df[['EUR','EE','ELPR','longitude','latitude']]
heat_maps = heat_maps.corr()

plt.figure(figsize=(10,6))
sns.set_context('notebook',font_scale=1)
sns.heatmap(heat_maps, annot=True,cmap='summer');
```

Department of Artificial Intelligence and Data Science



Department of Artificial Intelligence and Data Science



Methodological Details

- **Needs Assessment:** The program begins with a needs assessment to determine the participants' learning requirements and objectives. A pre-internship survey and individual interviews with participants are used to determine their expectations, aspirations, and past knowledge.
- **The theoretical lectures:** This lecture would discuss the theoretical framework that underpins the study, drawing on relevant theories and concepts from education, psychology, and statistics. This may include theories of learning, motivation, and achievement, as well as statistical models and methods for data analysis and prediction.
- **Practical tasks:** The next practical task would involve validating the model to ensure that it accurately predicts percentage scores based on study hours. This could involve using cross-validation, bootstrapping, or holdout validation techniques.
- **Hands-on Projects:** One hands-on project could involve creating visualizations of the data to identify any patterns or relationships between study hours and percentage scores. This could involve creating scatterplots, box plots, or other visualizations using tools such as R, Python, or Excel.
- **Mentorship & Support:** Students could be paired with a mentor who is experienced in data analysis, statistical modeling, or machine learning. The mentor could provide guidance on the research process, assist with the selection and interpretation of statistical models, and provide feedback on the students' work.
- **Assessment and Evaluation:** The program includes assessments and evaluations to track learning outcomes and internship effectiveness. Practical assignments, hands-on projects, and online quizzes are used to assess participants. The assessments are carried out using feedback questionnaires and participant interviews to identify the program's strengths and flaws as well as possibilities for development.

Result/Analysis/Conclusion

Result :

The recommendation system was integrated with the store's existing systems and processes, and the results were evaluated to determine the effectiveness of the recommendation engine. The evaluation included measuring the accuracy of the recommendations, the impact on sales, and customer satisfaction.

The results of the evaluation showed that the recommendation engine was able to generate personalized recommendations that were relevant and accurate. The accuracy of the recommendations was measured using metrics such as precision and recall, which showed that the recommendation engine was able to provide relevant recommendations for a high percentage of customer interactions..

Analysis :

The analysis of this data science project on product recommendations for retail stores provides insights into the potential benefits of leveraging customer data and machine learning algorithms to generate personalized recommendations. The project demonstrates that personalized recommendations can drive sales, improve customer satisfaction, and increase loyalty, highlighting the importance of providing a personalized shopping experience in the competitive retail landscape.

Conclusion :

Embarking on this Python project to analyze Iris unemployment data in India will not only enhance your technical skills but also enable you to make a meaningful impact on society. By uncovering the underlying causes, visualizing the data, and creating predictive models, you will contribute to empowering decision-makers and driving effective interventions to combat unemployment in India. Together, let's leverage the power of data and Python programming to inspire change, uplift communities, and create a brighter future for all.

Suggestions

- **Establish Specific Goals and Objectives:** Establish specific goals and objectives for the internship program and convey them to participants. This will assist students in understanding what is expected of them and what they may anticipate from the program.
- **Provide a Planned Curriculum:** Provide a structured curriculum covering the fundamentals of Data Science And Business Analytics. This ensures that participants have a firm foundation of knowledge and skills upon which to expand.
- **Encourage Collaboration:** Encourage people to work on projects and activities together. This will allow them to learn from one another, share ideas, and improve their collaboration abilities.
- **Provide Feedback Opportunities:** Give participants regular feedback on their progress and constructive criticism to help them develop their abilities. This will keep them motivated and focused on their objectives.
- **Mentorship:** Provide mentorship to participants from experienced individuals in the field of web development. This will give students with career counselling, assistance, and industry insights that will help them succeed.
- **Encourage Networking:** Encourage participants to network with other web development experts. Attending business events, participating in online forums, and engaging in social media activities are all ways to do this.

Acknowledgment

It is our pleasure to acknowledge a sense of gratitude to all those who helped us in the completion of the Internship. I am highly indebted to **Mrs. Anuja Garande** from **Zeal College, Narhe** & **Mr. Nilesh Kamble** for their guidance and constant supervision as well as for providing necessary information regarding the internship & also for their support.

I would like to express my gratitude towards **Mr. Dikshendra Sarpate (Head - Department of AIDS)** for their kind co-operation and encouragement which help me for providing the required facilities.

Finally, we wish to thank and appreciate all our teachers and friends for their constructive comments, suggestions, and guidance and all those who directly or indirectly helped us in completing this internship.

List of References

1. <https://www.python.org/downloads/>
2. <https://www.kaggle.com/datasets>
3. <https://pandas.pydata.org/>
4. <https://www.nvidia.com/en-us/glossary/data-science/recommendation-system>
5. <https://gloud.ai/blog/10-types-of-product-recommendations-for-shopify>
6. <https://wecanflyagency.com/blog/personalized-product-recommendations/>

Evaluation Sheet

Sr No.	Evaluation Parameters	Marks
1	Attendance Record (Out of 25)	
2	Diary/Workbook (Out of 25)	

Signature of Internal Internship Supervisor

Signature of External Internship Supervisor