INTERNSHIP REPORT

INTERNSHIP REPORT

On

DATA SCIENCE USING PYTHON

CHINTHAGUNTA VAMSHI KRISHNA
22951A66G2

DATA SCIENCE USING PYTHON

An INTERNSHIP REPORT

Bachelor of Technology
In
Computer Science and
Engineering

by

CHINTHAGUNTA VAMSHI KRISHNA
22951A66G2

Department of Artificial Intelligence and Machine Learning

INSTITUTE OF AERONAUTICAL ENGINEERING

(Autonomous)
Dundigal, Hyderabad – 500 043, Telangana
July, 2023

DECLARATION

I certify that

a. The work contained in this report is original and has been done by me under the guidance of my supervisor (s).

b. The work has not been submitted to any other Institute for any degree or diploma.

c. I have followed the guidelines provided by the Institute for preparing the report.

d. I have conformed to the norms and guidelines given in the Code of Conduct of the Institute.

e. Whenever I have used materials (data, theoretical analysis, figures, and text) from other sources, I have given due credit to them by citing them in the text of the report and giving their details in the references. Further, I have taken permission from the copyright owners of the sources, whenever necessary.

Ch. Vamshi Krishna

Place: Dundigal, Hyderabad

Signature of the Student

Date: 17/07/2023

22951A66G2

CERTIFICATE

राष्ट्रीय लघु उद्योग निगम–तकनीकी सेवा केन्द्र

THE NATIONAL SMALL INDUSTRIES CORPORATION LTD.

TECHNICAL SERVICES CENTRE

(भारत सरकार का उद्यम / A Government of India Enterprises)

ई.सी.आई.एल एक्स रोड, कुशाईगुडा, हैदराबाद – 500062, तेलंगाना, भारत
E.C.I.L X Road, Kushaiguda, Hyderabad - 500062, Telangana, India.

एन एस आई सी
N S I C

क्रमांक / S.No. 205352

दिनांक / Date: 31/07/2023

Certificate

This is to certify that Mr. / Ms. Chinthagunta Vamshi Krishna
son/daughter of Mr. Chinthagunta Suresh pursuing BTech in AI&ML from
(College Name) Institute of Aeronautical Engineering, Hyderabad
Roll No. 22951A66G2 has successfully completed the Field Project
entitled/in the area of Data Science Using Python
under
our guidance. It is a bonafide work carried out by her/him from 17/07/2023 to 31/07/2023
He/She has completed the assigned module as per the requirements within the time frame
During the above period, the trainee's conduct was found Good

Project Coordinator

Centre Head

यह प्रमाण पत्र होलोग्राम लगा होने से मान्य होगा / This Certificate shall be valid only with affixed hologram

APPROVAL SHEET

This **INTERNSHIP REPORT** entitled **Data Science Using Python** by **CHINTHAGUNTA VAMSHI KRISHNA** is approved for the award of the Degree Bachelor of Technology in **Computer Science and Engineering (AI & ML).**

Supervisor

Head of the Department

Head of the Department
Artificial Intelligence & Machine Learning
INSTITUTE OF AERONAUTICAL ENGINEERING
Dundigal, Hyderabad - 500 043

Date: 28/03/2024

Place: Hyderabad .

ABSTRACT

Keywords: NSIC, internship, data science, Python, data manipulation, exploratory data analysis, statistical analysis, machine learning algorithms, Pandas, NumPy, Matplotlib, Scikit-learn, Tensor Flow, practical application, real-world relevance.

The internship program conducted by the **National Small Industries Corporation (NSIC)** was meticulously designed to provide participants with a comprehensive understanding of data science principles and practical skills in Python programming. Over the duration of the program, participants were immersed in a structured curriculum that covered a broad spectrum of topics essential for data analysis and machine learning.

The program curriculum included in-depth sessions on data manipulation, exploratory data analysis (**EDA**), statistical analysis, and the application of various machine learning algorithms. Participants were introduced to a range of Python libraries such as **Pandas** for data manipulation, **NumPy** for numerical computing, **Matplotlib** for data visualization, and **Scikit-learn** for machine learning tasks.

A distinguishing feature of the internship program was its hands-on approach, which allowed participants to apply theoretical concepts directly to practical scenarios. Through a combination of theoretical lectures, hands-on exercises, and real-world case studies, participants gained proficiency in data handling, analysis, and interpretation

Furthermore, the internship fostered an environment of collaboration and learning, with participants encouraged to share insights and experiences. Expert guidance and mentorship were provided throughout the program to support participants in their learning journey.

Overall, the NSIC internship program on data science using Python provided participants with a solid foundation in data science methodologies and practical skills necessary to excel in the field. By the end of the program, participants were equipped with the knowledge and confidence to tackle complex data-driven problems and make informed decisions using data-driven insights.

Internship Report Brief Write up

Introduction:

In response to the burgeoning demand for skilled data professionals in today's data-driven world, the Institute of Aeronautical Engineering (IAE) undertook a proactive initiative to host an immersive internship program on data science using Python. This program, conducted in collaboration with the National Small Industries Corporation (NSIC), was meticulously designed to equip participants with practical skills and knowledge essential for navigating the complexities of modern data analysis and machine learning.

Curriculum Overview:

The internship program's curriculum was thoughtfully structured to cover a wide spectrum of topics essential for proficiency in data science using Python. Each component of the curriculum was carefully designed to build upon the foundational knowledge and gradually introduce more advanced concepts:

- **Foundations of Python**: Participants were introduced to the fundamentals of Python programming language, covering basic syntax, data structures, functions, and control flow.
- **Data Manipulation with Pandas:** Hands-on sessions focused on data manipulation and analysis using the Pandas library, encompassing data ingestion, cleaning, transformation, and aggregation.
- **Exploratory Data Analysis (EDA):** Techniques for exploring and visualizing data to uncover patterns, trends, outliers, and relationships, facilitating data-driven decision-making.
- **Statistical Analysis:** An overview of statistical methods and hypothesis testing, enabling participants to draw meaningful insights and conclusions from data.
- **Machine Learning Fundamentals:** Introduction to supervised and unsupervised learning algorithms, including regression, classification, clustering, and dimensionality reduction.

Methodology:

The internship program adopted a multifaceted methodology to cater to diverse learning styles and maximize participant engagement:

- **Theoretical Lectures:** Expert instructors delivered comprehensive lectures on theoretical concepts, providing participants with a solid theoretical foundation while emphasizing practical relevance and real-world applications.
- **Hands-on Workshops:** Interactive workshops provided participants with opportunities to apply theoretical concepts to real-world datasets and scenarios, reinforcing learning through practical experience.

- **Project-based Learning:** Participants engaged in industry-relevant projects and case studies, allowing them to apply their skills to solve complex data science problems and gain first-hand experience in addressing real-world challenges.

- **Peer Learning and Collaboration:** Opportunities for peer learning, collaboration, and knowledge sharing were fostered throughout the program, enabling participants to learn from each other's experiences, perspectives, and insights.

Key Learning Components:

Throughout the internship program, participants focused on developing a broad range of key competencies essential for success in data science roles:

- **Data wrangling and Pre-processing:** Skills in cleaning, pre-processing, and transforming raw data into a format suitable for analysis and modelling.
- **Data Visualization:** Proficiency in visualizing data using tools such as Matplotlib and Seaborn to communicate insights effectively and facilitate data-driven decision-making.
- **Statistical Analysis and Inference:** Understanding of fundamental statistical concepts and techniques for exploratory data analysis, hypothesis testing, and inference.
- **Machine Learning Modelling:** Exposure to a variety of machine learning algorithms and techniques for predictive modelling, classification, regression, and clustering.
- **Model Evaluation and Validation:** Techniques for evaluating model performance, assessing model robustness, and ensuring generalization to unseen data.
- **Model Deployment and Productionization:** Strategies for deploying machine learning models into production environments, including considerations for scalability, reliability, and maintainability.

Practical Application and Real-world Relevance:

A hallmark of the internship program was its emphasis on practical application and real-world relevance. Participants had the opportunity to work on a diverse range of industry-relevant projects and case studies, enabling them to apply their newly acquired skills to solve real-world problems across various domains such as finance, healthcare, marketing, and e-commerce. By engaging in hands-on projects, participants gained invaluable experience and insights into the practical applications of data science in different industries and contexts.

Challenges and Triumphs:

No journey is without its challenges, and mine was no exception. From debugging code to tackling complex algorithms, I faced obstacles head-on, fueled by determination and resilience. With each triumph over adversity, I emerged stronger, more confident, and more adept at navigating the intricacies of data science.

Outcome and Impact:

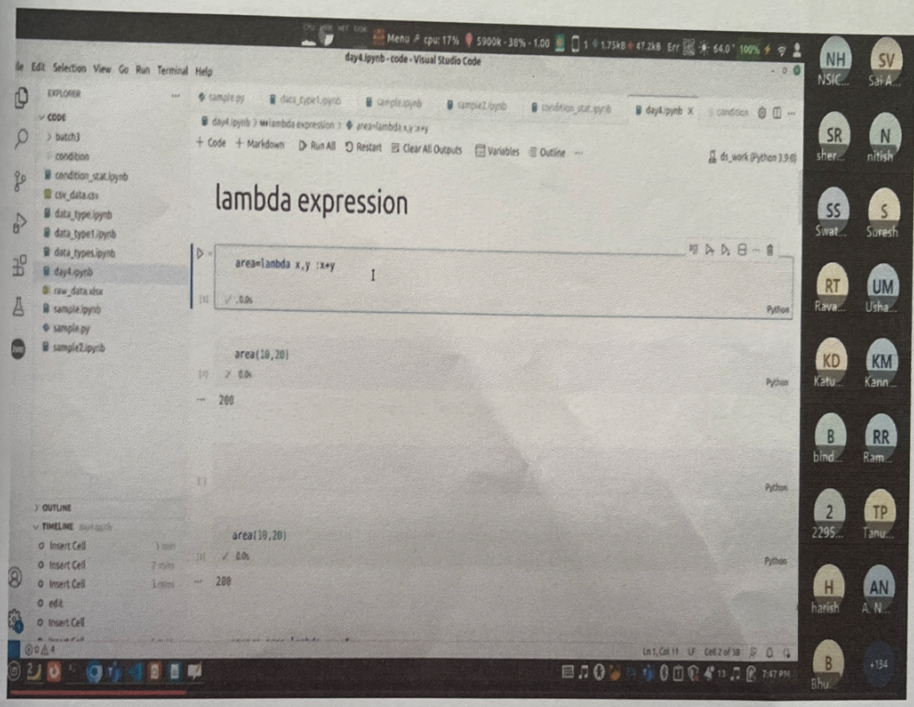
The internship program had a profound impact on participants, yielding significant outcomes in terms of skill enhancement, professional development, and career advancement:

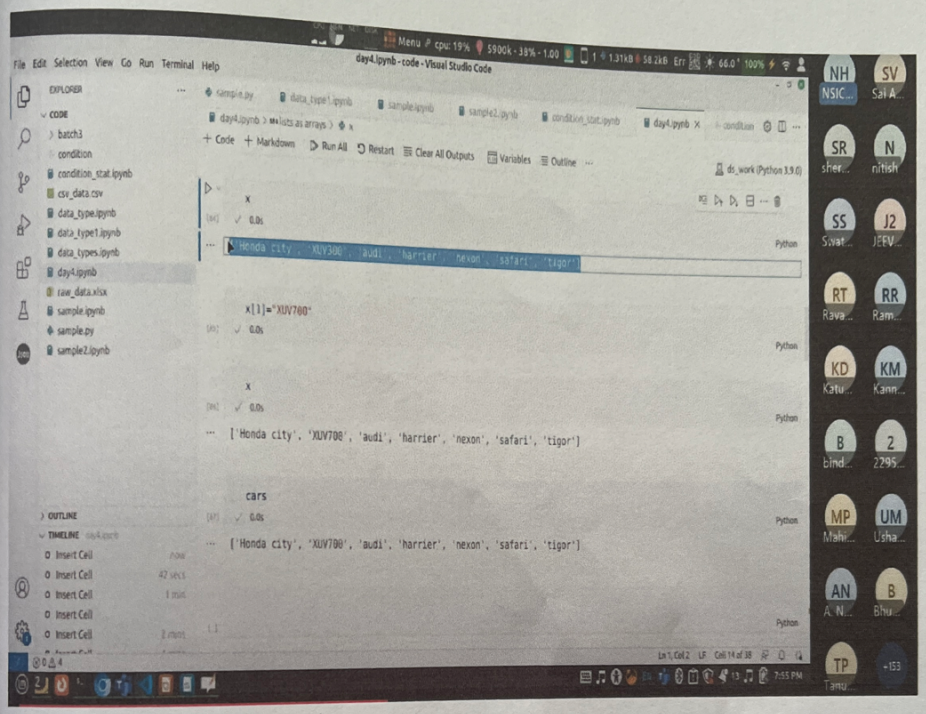
- **Enhanced Technical Proficiency:** Participants emerged from the program with enhanced technical skills and proficiency in Python programming, data manipulation, exploratory data analysis, statistical analysis, and machine learning modelling.
- **Increased Confidence:** The hands-on nature of the program and exposure to real-world projects boosted participants' confidence in their ability to tackle complex data science problems and make data-driven decisions.
- **Expanded Career Opportunities:** The practical skills and industry-relevant experience gained during the internship program enhanced participants' employability and opened up new career opportunities in the rapidly growing field of data science.
- **Professional Networking:** Participants had the opportunity to network with industry professionals, experts, and peers, expanding their professional network and fostering potential collaborations, mentorships, and career opportunities.

Conclusion:

In conclusion, the internship program on data science using Python hosted by the Institute of Aeronautical Engineering and presented by NSIC was a resounding success, providing participants with a transformative learning experience and equipping them with the skills, knowledge, and confidence needed to excel in the dynamic field of data science. By combining theoretical learning with practical application, the program prepared participants to tackle real-world data challenges, drive data-driven decision-making, and make meaningful contributions to their organizations and the broader data science community. As data continues to play an increasingly pivotal role in driving innovation, informing decision-making, and solving complex problems across industries, the skills and experiences gained through this internship program will undoubtedly position participants for success in their future endeavours.

Screen shots of Virtual Training.




day4.ipynb - code - Visual Studio Code
File Edit Selection View Go Run Terminal Help
EXPLORER
CODE
batch3
condition
condition_stat.ipynb
csv_data.csv
data_type.ipynb
data_type1.ipynb
data_types.ipynb
day4.ipynb
raw_data.xlsx
sample.ipynb
sample.py
sample2.ipynb
OUTLINE
TIMELINE
+ Code + Markdown Run All Restart Clear All Outputs Variables Outline
x
['Honda city', 'XUV300', 'audi', 'harrier', 'nexon', 'safari', 'tigor']
x[1]='XUV700'
['Honda city', 'XUV700', 'audi', 'harrier', 'nexon', 'safari', 'tigor']
cars
['Honda city', 'XUV700', 'audi', 'harrier', 'nexon', 'safari', 'tigor']

THANK YOU