

**Course Handout**

Institute/School Name	Chitkara University Institute of Engineering & Technology		
Department Name	Department of Computer Science & Engineering		
Programme Name	Bachelor of Engineering- Computer Science & Engineering (Artificial Intelligence and Machine Learning)		
Course Name	Data Analytics	Session	July-Dec, 2025
Course Code	24CAI0107	Semester/Batch	3 <sup>rd</sup> /2024
L-T-P (Per Week)	2-0-2	Course Credits	3
Pre-requisite	Knowledge of statistics, programming (preferably Python)	NHEQF Level	5
Course Coordinator	Dr. Harjeet Singh	SDG Number	4, 9

**1. Objectives of the Course**

With great amounts of data comes a great need for data analysts. Organizations generate and collect an exponentially growing amount of data: wringing actionable answers and insights out of the chaos is a valuable and in-demand skill set to have B2B and B2C commerce, health care, manufacturing, and marketing all use data analytics to improve processes and enhance profits. This course takes students through five key factors that indicate the need for specific AWS services in collecting, processing, analysing, and presenting your data. This includes learning basic architecture, value propositions, and potential use cases. The core objectives of this course are:

- To gain an immersive understanding of the practices and processes used by a data analyst in their day-to-day job operations.
- To understand data cleaning and organizing for analysis, and calculations using spreadsheets, SQL and Python programming.
- To inculcate key analytical skills and understanding of tools like Python programming, Tableau, Power BI.
- To implement visualization and presentation of data findings in dashboards, and commonly used visualization platforms.

**2. Course Learning Outcomes (CLOs)**

Students should be able to:

Course Learning Outcomes	CLOs	Program Outcomes (PO)	NHEQF Level Descriptor	No. of Lectures
<b>CLO1</b>	Develop the ability to build proficiency in statistical analysis of data.	PO1, PO4	Q1, Q2	4
<b>CLO2</b>	Apply data science concepts and methods to solve problems in real-world contexts and communicate these solutions effectively.	PO10, PO11	Q3, Q6	20
<b>CLO3</b>	Understand data visualization techniques to present and interpret data insights effectively.	PO5, PO10	Q2, Q3	10

<b>CLO4</b>	Perform data cleaning and transform variables to facilitate analysis by integrating data from disparate sources.	PO1, PO2	Q1, Q2	10
<b>CLO5</b>	Build and enhance business intelligence capabilities by adapting the appropriate technology and software solutions.	PO7, PO10	Q6	16
<b>Total Contact Hours</b>				<b>60</b>

**CLO-PO-PSO Mapping grid |Program Outcomes (POs) and Program Specific Outcomes (PSO) are available as a part of Academic Program Guide**

CLO	P01	P02	P03	P04	P05	P06	P07	P08	P09	P10	P11	P12	PS01	PS02	PS03	Type of Assessment
CLO01	L	L	L	L	L					L	L	M	M	L	M	Formative /Summative
CLO02	H	H	H	H	M					L	L	M	M	M	H	Formative /Summative
CLO03	H	H	H	H	M					L	L	M	M	H	M	Formative /Summative
CLO04	H	H	H	H	M					L	L	M	H	M	M	Formative /Summative
CLO05	M	M	M	H	M					L	L	M	L	L	L	Formative /Summative

\*H=High, M=Medium, L=Low

### 3. Recommended Books:

- B01:** JojoMoolayil, “Smarter Decisions: The Intersection of IoT and Data Science”, PACKT.  
**B02:** Cathy O’Neil and Rachel Schutt, “Doing Data Science”, O’Reilly.  
**B03:** David Dietrich, Barry Heller, Beibei Yang, “Data Science and Big data Analytics”, EMC.  
**B04:** Raj, Pethuru, “Handbook of Research on Cloud Infrastructures for Big Data Analytics”, IGI Global.  
**B05:** Greg Deckler, Bret Powell, “Mastering Microsoft Power BI: Expert Techniques to Create Interactive, Intelligence”, Packt Publishing, 2<sup>nd</sup> Edition.  
**B06:** Angela Dean, Daniel Voss, and Danel Draguljić, “Insights for Effective Data Analytics and Business Design and Analysis of Experiments”.

### 4. Other readings and relevant websites:

Serial No.	Link of Journals, Magazines, websites and Research Papers
1.	<a href="https://www.tutorialspoint.com/statistics/index.html">https://www.tutorialspoint.com/statistics/index.html</a>
2.	<a href="https://iridl.ldeo.columbia.edu/dochelp/StatTutorial/index.html">https://iridl.ldeo.columbia.edu/dochelp/StatTutorial/index.html</a>
3.	<a href="https://www.khanacademy.org/math/statistics-probability">https://www.khanacademy.org/math/statistics-probability</a>
4.	<a href="https://www.analyticsvidhya.com/blog/2021/02/an-intuitive-guide-to-visualization-in-python/">https://www.analyticsvidhya.com/blog/2021/02/an-intuitive-guide-to-visualization-in-python/</a>
5.	<a href="https://www.educba.com/data-science/data-science-tutorials/tableau-tutorial/">https://www.educba.com/data-science/data-science-tutorials/tableau-tutorial/</a>
6.	<a href="https://matplotlib.org/stable/tutorials/index.html">https://matplotlib.org/stable/tutorials/index.html</a>
7.	<a href="https://seaborn.pydata.org/tutorial.html">https://seaborn.pydata.org/tutorial.html</a>

### 5. Recommended Tools and Platforms

Python, Jupyter Notebook, Visual Studio Code, Anaconda, Tableau, Microsoft PowerBI

## 6. Course Plan: Theory+ Lab

### a. Theory Plan

Lect. No.	Topics
1	Understanding data: Introduction – Types of Data: Numeric – Categorical, Graphical, High Dimensional Data
2	Classification of digital Data: Structured, Semi-Structured and Un-Structured, Sources of Data
3	Case Studies on Different types of Data: Time Series, Transactional Data, Biological Data, Spatial Data, Social Network Data
4	Data Evolution: Understand issues relating to acquisition, cleaning and loading of data, Data Deluge, Data lake
5-7	Python Numpy: Arrays, Indexing, Slicing; Different Array Operations, Linear Algebra Operations
8-12	Python Pandas: Dataframes, Handling of data Series, Data wrangling, Alignment and Indexing, Handling Missing Data, Data Cleaning, Merging and Joining Dataframes, Grouping, Concatenation and Aggregation, Masking, Performing Mathematical Operations on Data
13-16	Access and combine data from CSV, JSON, logs, APIs, and databases: Using Pandas to access different sources of Data
17-19	Using SQL with Databases: DDL, DML, Select and Joins
20-23	Advanced Operations using Pandas: Statistical Functions, Descriptive Statistics, Working with Text Data, Time Delta. Basic Data Visualization using pandas
<b>Sessional Test-1 (1-23 Lectures)</b>	
24-28	Matplotlib, Seaborn, Cufflinks: Different Plot types, Scatter, bar, histogram, box, pie, violin, Subplots, axis and figures, Text, labels and annotations, Colormaps, Plotting with Seaborn, Plotting Categorical and Continuous Data, Visualizing Regression models, Plotting interactive plots using Cufflinks
29-33	Using Power BI for Visualization: Connecting to different data sources, Data types, Working with Meta Data, Calculations, Purpose of Data Analysis Expressions (DAX), DAX operators & Functions, Power Query, Different Charts& Reports, Exploring Data geographically, Building dashboard to see insights
34-38	Exploratory data analysis: Missing value analysis, Inferential Statistics: Different Data distributions
39-44	Regular Expressions: Matching Exact Words Special Characters and Metacharacters Using Character Classes and Ranges Anchors and Boundaries Quantifiers in Regex Escaping Special Characters Extracting Email IDs and Phone Numbers Validating Passwords or Usernames Cleaning and Extracting from URLs Detecting Dates and Times in Text Replacing Text Patterns Splitting Strings Using Regex
<b>Sessional Test-2 (24-44 Lectures)</b>	
45-49	Web Scraping Use Case and Legality, Requests and BeautifulSoup, HTML and CSS Basics, Extracting Text, Links and Images, Navigating HTML Structures
50-52	Understanding Regression: Linear and Non-linear Regression

52-55	Identification of regression problems: Case Studies
56-60	Identifications of Classification problems: Case Studies
<b>END-TERM EXAM (FULL SYLLABUS)</b>	

**b. Lab Plan:**

Lab No.	Experiment
1	Create and manipulate NumPy arrays and Pandas DataFrames to perform basic data operations.
2	Identify and handle missing data in datasets using Pandas, including techniques like imputation and removal.
3	Perform data wrangling tasks such as aligning data, setting indexes, merging, and joining multiple DataFrames.
4	Implement Pandas to group data, apply masking techniques, and perform mathematical operations to derive insights.
5	Read data from various sources like CSV, JSON, logs, APIs, and databases, and combine them into a unified DataFrame.
6-7	Create various plots and visualizations using Pandas, Matplotlib, and Seaborn to represent data insights effectively.
8-9	Utilize PowerBI to connect to various data sources and create interactive visualizations and dashboards.
10	Explore statistical concepts such as measures of central tendency, variability, skewness, and kurtosis using Python.
11	Conduct exploratory data analysis by handling missing values, creating correlation matrices, and detecting outliers.
12-13	Use Power Query to clean and transform raw data by handling missing values, filtering, merging, and splitting columns.
14-15	Use Power BI's visualization tools to create bar charts, line graphs, scatter plots, and pie charts to analyze trends.
16	Apply filters and slicers to Power BI reports for interactive data exploration and drill-down analysis.
17	Write DAX expressions to create calculated columns, measures, and aggregations like SUM, AVERAGE, and COUNT.
18	Design and build an interactive Power BI dashboard with multiple visualizations, KPIs, and real-time updates.
<b>Lab Evaluation (1-18 Lab No.)</b>	

**7. Delivery/Instructional Resources**

**Plan (Theory +Lab):**

Lec. No.	Topic	CLO	Book No, CH No, Page No	TLM	ALM	Web References	Audio-Video
1	Understanding data	CLO01	B03, CH No. 2	Lecture, Discussion	Think/pair/share	<a href="https://www.sqlshack.com/introduction-to-data-science-data-understanding-and-preparation/">https://www.sqlshack.com/introduction-to-data-science-data-understanding-and-preparation/</a> <a href="https://ibm-cloud-architecture.github.io/refarch-data-ai-analytics/preparation/data-understanding/">https://ibm-cloud-architecture.github.io/refarch-data-ai-analytics/preparation/data-understanding/</a>	
2	Classification of digital Data	CLO01	B03, Ch No. 3	Lecture, Discussion	Quiz/ Test Questions	<a href="https://ibm-cloud-architecture.github.io/refarch-data-ai-analytics/preparation/data-understanding/">https://ibm-cloud-architecture.github.io/refarch-data-ai-analytics/preparation/data-understanding/</a>	<a href="https://www.youtube.com/watch?v=mm2A5tKVIpg">https://www.youtube.com/watch?v=mm2A5tKVIpg</a>
3	Case studies on types of data	CLO01	B02, Ch no. 1	Lecture, Discussion	Leading Questions	<a href="https://data-flair.training/blogs/big-data-case-studies/">https://data-flair.training/blogs/big-data-case-studies/</a>	
4	Data Evolution	CLO02	B02, Ch no. 5	Lecture, Demonstration	Quiz, Peer Reviews	<a href="https://www.kdnuggets.com/2014/06/data-lakes-vs-data-warehouses.html">https://www.kdnuggets.com/2014/06/data-lakes-vs-data-warehouses.html</a>	<a href="https://www.youtube.com/watch?v=E49BFhThC3U">https://www.youtube.com/watch?v=E49BFhThC3U</a>
5-7	Python Numpy	CLO02	B01, Ch no. 1 and 2	Lecture, Questioning	Test questions/ Quiz	<a href="https://numpy.org/doc/stable/user/index.html#user">https://numpy.org/doc/stable/user/index.html#user</a>	<a href="https://www.youtube.com/watch?v=j31ah5Qa4QI">https://www.youtube.com/watch?v=j31ah5Qa4QI</a>
8-12	Python Pandas	CLO03	B01, Ch no. 5	Lecture, Demonstration	Brainstorming Session	<a href="https://pandas.pydata.org/docs/user_guide/index.html">https://pandas.pydata.org/docs/user_guide/index.html</a>	<a href="https://www.youtube.com/watch?v=UB3DE5Bgfx4">https://www.youtube.com/watch?v=UB3DE5Bgfx4</a>
13-16	Access and combine data from CSV, JSON, logs, APIs, and databases	CLO03	B01, Ch no. 5	Lecture, Discussion	Test Questions	<a href="https://pandas.pydata.org/docs/user_guide/io.html">https://pandas.pydata.org/docs/user_guide/io.html</a>	<a href="https://www.youtube.com/watch?v=GFBxxxjAzaU">https://www.youtube.com/watch?v=GFBxxxjAzaU</a>
17-19	Using SQL with Databases	CLO03	B01, Ch no. 7, 8 and 9	Lecture, Questioning	Brainstorming Session	<a href="https://www.sqltutorial.org/">https://www.sqltutorial.org/</a>	<a href="https://www.youtube.com/watch?v=zbMHLJ0dY4w">https://www.youtube.com/watch?v=zbMHLJ0dY4w</a>
20-23	Advanced Operations using Pandas	CLO05	B02, Ch no. 9	Lecture, Reviewing	Test Questions	<a href="https://www.kdnuggets.com/2019/10/5-advanced-features-">https://www.kdnuggets.com/2019/10/5-advanced-features-</a>	<a href="https://www.youtube.com/watch?v=DUGd48QYmfl">https://www.youtube.com/watch?v=DUGd48QYmfl</a> <a href="https://www.youtube.com/watch?v=DUGd48QYmfl">https://www.youtube.com/watch?v=DUGd48QYmfl</a>

						<a href="#">pandas.html</a>	<a href="#">utube.com/watch?v=RIliVeig3hc</a>
24-28	Matplotlib, Seaborn, Cufflinks	CLO05	B04, Ch no. 1, 2, 3 and 4	Lecture	Leading Questions	<a href="https://matplotlib.org/stable/plot_types/index.html">https://matplotlib.org/stable/plot_types/index.html</a> <a href="https://seaborn.pydata.org/tutorial.html">https://seaborn.pydata.org/tutorial.html</a> <a href="https://www.analyticsvidhya.com/blog/2021/06/advanced-python-data-visualization-libraries-plotly/">https://www.analyticsvidhya.com/blog/2021/06/advanced-python-data-visualization-libraries-plotly/</a>	<a href="https://www.youtube.com/watch?v=3Xc3CA655Y4">https://www.youtube.com/watch?v=3Xc3CA655Y4</a> <a href="https://www.youtube.com/watch?v=ooqXQ37XHMM">https://www.youtube.com/watch?v=ooqXQ37XHMM</a> <a href="https://www.youtube.com/watch?v=7n5GzKuvPsw">https://www.youtube.com/watch?v=7n5GzKuvPsw</a>
29-33	Using Power BI for Visualization	CLO03	B03	Discussion	Leading Questions	<a href="https://www.tutorialspoint.com/power_bi/index.htm">https://www.tutorialspoint.com/power_bi/index.htm</a>	<a href="https://www.youtube.com/watch?v=NalazxB0-90">https://www.youtube.com/watch?v=NalazxB0-90</a>
34-38	Exploratory data analysis	CLO04	B05	Demonstration	Brainstorming Session	<a href="https://realpython.com/python-statistics/">https://realpython.com/python-statistics/</a>	<a href="https://www.youtube.com/watch?v=mQ-3KwrBIN0">https://www.youtube.com/watch?v=mQ-3KwrBIN0</a>
39-44	Regular Expressions	CLO04	B02	Lecture	Leading Questions	<a href="https://www.freecodecamp.org/news/how-to-scrape-websites-with-python-and-beautifulsoup-5946935d93fe/">https://www.freecodecamp.org/news/how-to-scrape-websites-with-python-and-beautifulsoup-5946935d93fe/</a>	<a href="https://youtu.be/mlHrfpkW-9o">https://youtu.be/mlHrfpkW-9o</a>
45-49	Web Scraping Use Case and Legality	CLO02	B04	Lecture, Discussion	Brainstorming Session	<a href="https://arxiv.org/abs/2009.10862">https://arxiv.org/abs/2009.10862</a>	<a href="https://glaspt.co/youtube/V59bYflomVk?utm_source=chatgpt.com">https://glaspt.co/youtube/V59bYflomVk?utm_source=chatgpt.com</a>
50-52	Understanding Regression	CLO05	B05	Lecture, Discussion	Brainstorming Session		<a href="https://glaspt.co/youtube/V59bYflomVk?utm_source=chatgpt.com">https://glaspt.co/youtube/V59bYflomVk?utm_source=chatgpt.com</a>
52-55	Identification of regression problems	CLO05	B03	Lecture, Questioning	Brainstorming Session	<a href="https://www.simplilearn.com/tutorials/data-analytics-tutorial/classification-vs-clustering">https://www.simplilearn.com/tutorials/data-analytics-tutorial/classification-vs-clustering</a>	<a href="https://www.youtube.com/watch?v=1R7-1hhhR1M">https://www.youtube.com/watch?v=1R7-1hhhR1M</a> <a href="#">reddit.com</a>
56-60	Identifications of Classification problems	CLO05	B04	Lecture	Brainstorming Session		<a href="https://www.youtube.com/watch?v=s6PSSzeUMFk">https://www.youtube.com/watch?v=s6PSSzeUMFk</a>

### 8. Remedial Classes

After every Sessional Test, different types of learners will be identified, and special discussions will be planned and scheduled accordingly for the slow learners.

### 9. Self-Learning

Assignments to promote self-learning, survey of contents from multiple sources.

S. No.	Topics	CLO	ALM	References/MOOCs
1	Pandas, Numpy, Seaborn	CLO02, CLO05	Think/pair/share	<a href="https://www.kaggle.com/learn/pandas">https://www.kaggle.com/learn/pandas</a> <a href="https://www.w3schools.com/python/numpy/default.asp">https://www.w3schools.com/python/numpy/default.asp</a>
2	PowerBI Dashboards	CLO01, CLO04, CLO05	Think/pair/share	<a href="https://learn.microsoft.com/en-us/training/powerplatform/power-bi/">https://learn.microsoft.com/en-us/training/powerplatform/power-bi/</a>

### 10. Delivery Details of Content Beyond Syllabus

Content beyond the syllabus covered (if any) should be delivered to all students that would be planned, and schedule notified accordingly.

S. No.	Advanced Topics, Additional Reading, Research papers and any	CLO	POs	ALM	References/MOOCs
1	Regression	CLO01, CLO02, CLO05	PO1, PO2, PO3, PO4, PO7, PO11	Think/pair/share	<a href="https://scikit-learn.org/stable/auto_examples/linear_model/plot_ols.html">https://scikit-learn.org/stable/auto_examples/linear_model/plot_ols.html</a>
2	Classification and Clustering	CLO01, CLO03, CLO05	PO1, PO2, PO3, PO5, PO7, PO11	Think/pair/share	<a href="https://scikit-learn.org/stable/supervised_learning.html#supervised-learning">https://scikit-learn.org/stable/supervised_learning.html#supervised-learning</a>

### 11. Evaluation Scheme & Components:

Assessment Type	Evaluation Component	Type of Component	No. of Assessments	% Weightage of Component	Max. Marks	Mode of Assessment	CLO
Formative	Component1	Continuous Lab Evaluation	01*	20%	20	Offline	CLO01, CLO02, CLO03, CLO04, CLO05
Summative	Component 2	Sessional Tests (STs)	02**	30%	30	Offline	CLO01, CLO02, CLO03, CLO04, CLO05
Summative	Component 3	End Term Examination	01***	50%	50	Offline	CLO01, CLO02, CLO03, CLO04,

							CLO05
<b>Total</b>				<b>100%</b>			

\* Lab Evaluation is mandatory for all the students.

\*\* All STs are mandatory. The average of both STs will be taken for internal assessment.

\*\*\* To appear for the End Term Exam, attendance must be at least 75%.

## 12. Syllabus of the Course:

<b>Subject:</b> Data Analytics		<b>Course Code:</b> 24CAI0107	
<b>Lec. No.</b>	<b>Topic (s)</b>	<b>No. of Lectures</b>	<b>Weightage %</b>
1-4	Understanding data: Introduction, Types of Data: Numeric, Categorical, Graphical, High Dimensional Data, Classification of digital Data: Structured, Semi-Structured and Un-Structured, Sources of Data, Case Studies on Different types of Data, Time Series, Transactional Data, Biological Data, Spatial Data, Social Network Data, Data Evolution, Understand issues relating to acquisition, cleaning and loading of data, Data Deluge, Data lake.	4	10%
5-19	Python Numpy, Python Pandas: Dataframes, Handling of data, Data wrangling, Alignment and Indexing, Handling Missing Data, Data Cleaning, Merging and Joining Dataframes, Grouping, Concatenation and Aggregation, Masking, Performing Mathematical Operations on Data, Access and combine data from CSV, JSON, logs, APIs, and databases, Using SQL with Databases: DDL, DML, Select and Joins.	15	25%
20-33	Advanced Operations using Pandas: Statistical Functions, Descriptive Statistics, Working with Text Data, Time Delta, Basic Data Visualization using pandas, Matplotlib, Seaborn, Cufflinks, Using PowerBI for Visualization, Connecting to different data sources, Data types, Working with Meta Data, Calculations, Purpose of Data Analysis Expressions (DAX), DAX operators & Functions, Power Query, Different Charts & Reports, Exploring Data geographically, Building dashboard to see insights	14	20%
34-38	Exploratory data analysis: Missing value analysis, Inferential Statistics: Different Data distributions	5	8%
39-44	Regular Expressions: Matching Exact Words Special Characters and Metacharacters Using Character Classes and Ranges Anchors and Boundaries Quantifiers in Regex Escaping Special Characters Extracting Email IDs and Phone Numbers Validating Passwords or Usernames Cleaning and Extracting from URLs Detecting Dates and Times in Text Replacing Text Patterns Splitting Strings Using Regex	6	10%
45-49	Web Scraping Use Case and Legality, Requests and BeautifulSoup, HTML and CSS Basics, Extracting Text, Links and Images, Navigating HTML Structures	5	9%
50-60	Understanding Regression: Linear and Non-linear Regression, Identification of regression and classification problems: Case Studies, Identifications of Classification problems: Case Studies	11	18%

## 13. Academic Integrity Policy:

Education at Chitkara University builds on the principle that excellence requires freedom where Honesty and



integrity are its prerequisites. Academic honesty in the advancement of knowledge requires that all students and Faculty respect the integrity of one another's work and recognize the importance of acknowledging and safeguarding intellectual property. Any breach of the same will be tantamount to severe academic penalties.

**This Document is approved by:**

Designation	Name	Signature
Course Coordinator	Dr. Harjeet Singh	
Head-Academic Delivery	Dr. Kamal Deep Garg	
Dean (CSE-AI)	Dr. Sushil Kumar Narang	
Date (DD/MM/YYYY)	27/06/2025	