



## **Student Guide for Sync Session**

### **<<Week 17: Exploratory Data Analysis>>**

This guide is your roadmap to making the most of our online session. Packed with essential tips and strategies, it's designed to keep you engaged, prepared, and ready to dive into a smooth and productive learning journey. Get ready to participate, learn, and thrive!

### **Session Overview**

<b>Session title</b>	Exploratory Data Analysis (EDA)	
<b>Session duration</b>	3 hours	
<b>Session type</b>	<ul style="list-style-type: none"><li><b>Lectures:</b> Foundational understanding and hands-on demonstration of EDA techniques using Python libraries.</li><li><b>Case Studies:</b> Exploratory analysis on a Used Car Prices dataset to derive insights and perform data preparation.</li></ul>	
<b>Scope</b>	<p>This session introduces participants to the core concepts and techniques of Exploratory Data Analysis (EDA). It covers:</p> <ul style="list-style-type: none"><li>The role of EDA in the data science pipeline.</li><li>Techniques to summarise and visualise datasets.</li><li>Identifying patterns, trends, and anomalies.</li><li>Performing hands-on EDA using Python libraries: Pandas, NumPy, Matplotlib, and Seaborn.</li></ul>	
<b>Learning Objectives</b>	<b>Objective</b>	<b>Core Capability</b>
	Understand EDA's role in data science workflows	Analytical thinking and data preparation
	Handle missing values and perform grouping operations	Skill in descriptive statistics and visual analysis
	Engineer new features and encode variables	Data cleaning and transformation skills
	Use summary statistics and visual tools to explore data	Insight extraction and pattern recognition
	Perform correlation analysis and univariate/bivariate visualisations	Feature engineering and data transformation
	Apply EDA best practices and communicate insights	Business problem-solving and stakeholder communication
<b>Software/Tools</b>	<ul style="list-style-type: none"><li>Python (Pandas, NumPy, Matplotlib, Seaborn)</li><li>IDE: Jupyter Notebook</li><li>Dataset: Used Car Prices Dataset (used_cars.csv)</li></ul>	



- Presentation Tool: PowerPoint

### Pro tips for success

- **Ask Bold Questions:** No question is too small—curiosity is the key to learning!
- **Be Hands-On:** Coding is your superpower. Tweak, test, and break things (safely) to learn.
- **Collaborate:** Share your ideas in discussions. You might just spark the next big insight!

### Session Details

Topic	A glimpse	Insight / Actionable
Introduction	Get oriented with the session's goals and objectives.	Think about how data-driven decisions shape our daily lives.
What is EDA?	Understand the purpose and value of EDA.	Consider what insights you'd want before building a model.
Loading and Previewing Data	Load data using Pandas and inspect its structure.	Use <code>df.head()</code> and <code>df.info()</code> to understand data.
Summary Statistics	Use <code>describe()</code> to summarise data.	Analyse central tendencies and variability.
Handling Missing Values	Detect and fill missing values.	Use <code>isnull()</code> and <code>fillna()</code> to clean your data.
Categorical Variable Analysis	Explore categorical variables.	Use <code>value_counts()</code> to summarise category frequency.
Grouping and Summarising	Aggregate data with <code>groupby()</code> .	Compare fuel type vs. price.
Correlation Analysis	Check relationships between variables.	Use <code>.corr()</code> and heatmaps to identify patterns.
Univariate Distribution	Visualise single variable distributions.	Use <code>sns.histplot()</code> and observe shapes.
Boxplot for Outliers	Detect price variation by category.	Use <code>boxplot()</code> to identify outliers.
Feature Creation	Create new variables.	Derive <code>car_age</code> , <code>price_per_km</code> .
Feature Engineering: Binning	Categorise continuous variables.	Use <code>pd.cut()</code> to group age.
Feature Engineering: Encoding	Encode categorical variables.	Use <code>pd.get_dummies()</code> for one-hot encoding.



## Advanced Certification Programme in Data Science and Business Analytics with Generative AI

EDA Best Practices	Summarise steps and avoid pitfalls.	Keep exploration structured and insightful.
Q&A	Clarify doubts and share insights.	Reflect on discoveries and applications.

### Post-Session Activities

<b>Reflection Challenge</b>	What feature did you engineer today that uncovered a new insight? How might it help build a better model?
<b>Explore More</b>	<ul style="list-style-type: none"><li><b>Read:</b> Seaborn documentation for data visualisation.</li><li><b>Watch:</b> YouTube tutorials on hands-on EDA.</li><li><b>Try:</b> Kaggle notebooks that show creative EDA workflows.</li></ul>
<b>Get Inspired</b>	Did you know? Many great ML models are built on strong EDA foundations. Netflix, Uber, and Airbnb all rely on it to spot trends and make strategic decisions.
<b>The Journey Ahead</b>	Practice EDA on a new dataset. Try applying techniques like feature creation, binning, and encoding to datasets on housing, health, or e-commerce.