Introduction + EDA - Methodologies and Best Practices

By this session, the learners would have been through Python for Data Science, Visualisation, and Exploratory Data Analysis (EDA). They would have learnt about the data science applications of Python using libraries such as numpy, pandas, matplotlib, and seaborn. Currently, they will have gone through their first assignment, which will involve extensive use of exploratory data analysis concepts and methodologies. Hence, it is important that the learners be well versed with it.

Learning Objective(s)

This is the first Small Group Coaching session that the students will be attending. Here the primary objective of the session is to interact with the learners and establish a rapport for the upcoming learning & development at upGrad.

Since EDA is an extremely broad topic, it is crucial that the students are guided well on it and are provided with ample support, knowledge, and methodologies related to EDA so that they are well prepared for the ongoing case study and all the analysis work that is to come in the future. The learners should be able to derive insights from raw data without any concrete questions being asked.

Prerequisite(s)

The learners are expected to have gone through:

- Python for Data Science,
- Data Visualisation in Python, and
- Exploratory Data Analysis.

Brief Agenda

1. Part - I: Introduction (25 min)
2. Part - II: Doubt resolution + P2P (10 min)
3. Part - II: Focused teaching (45 min)
4. Part - III: Doubt resolution (10 min)

Detailed Lesson Plan

Component	Instruction Task/Learner Task	Time (min)	# Questions	Element of Engagement
Part - I: Introduction (25 min)	The coach introduces him/ herself with professional details like: Work experience, where he/ she has worked all the years. Where is he/ she from and what skills are his/ her best? Then he/ she can introduce their hobbies, motivation behind teaching sessions on weekends and what are their expectations from the class. Once he has introduced himself, the coach will ask each learner to introduce themselves with : Name Background The coach then briefs learners about what is SGC and what can learners expect out of it through the slides shared by upgrad.	25		Social support
Part - II: Doubt Resolution (20 min)	The instructor will ask the learners to summarise and briefly explain their learning from the past few weeks (one on one).	10	-	Social support

	The instructor clears the doubts that the learners might have (learners will be asked to come up with a list of doubts, which will be made available to the instructor beforehand). *Both the instructor and the learners are advised to not spend too much time discussing generic and vague doubts.*		-	Doubt resolution
Part - III: Focused Teaching (40 min)	Exploratory data analysis is a broad topic. There are certain common practices that are followed here as well, although sometimes, it also tends to become a bit subjective depending upon the data or the business problem at hand. Hence, the focus of this SGC will not be to teach the learners hardcore EDA, but to focus more on task support, i.e., the instructor will be providing the students with ample support on the methodology and best practices to follow while performing EDA on any data. These practices include (not exhaustive) checking the head of the data after every major step, understanding when to replace entries with NA and when to delete/ignore them, identifying which columns should be checked for sanity, understanding how to leverage domain knowledge while selecting features for analysis, understanding how to choose a particular type of visualisation, and so on.	10	-	Personalised feedback
Part - IV: Doubt Resolution (15 min)	The instructor will address some of the most common doubts from Exploratory Data Analysis. Also, they should encourage the learners to try answering a few of them first before providing them with the	10	5	Doubt resolution

	necessary explanation. *Both the coach and the learners are advised to not spend too much time discussing generic and vague doubts.*			

Additional Reading for the Week

1. **Link 1**

   3 best practices for exploratory data visualizations

2. **Link 2**

   Why EDA is Crucial for Any Data Science Project?