

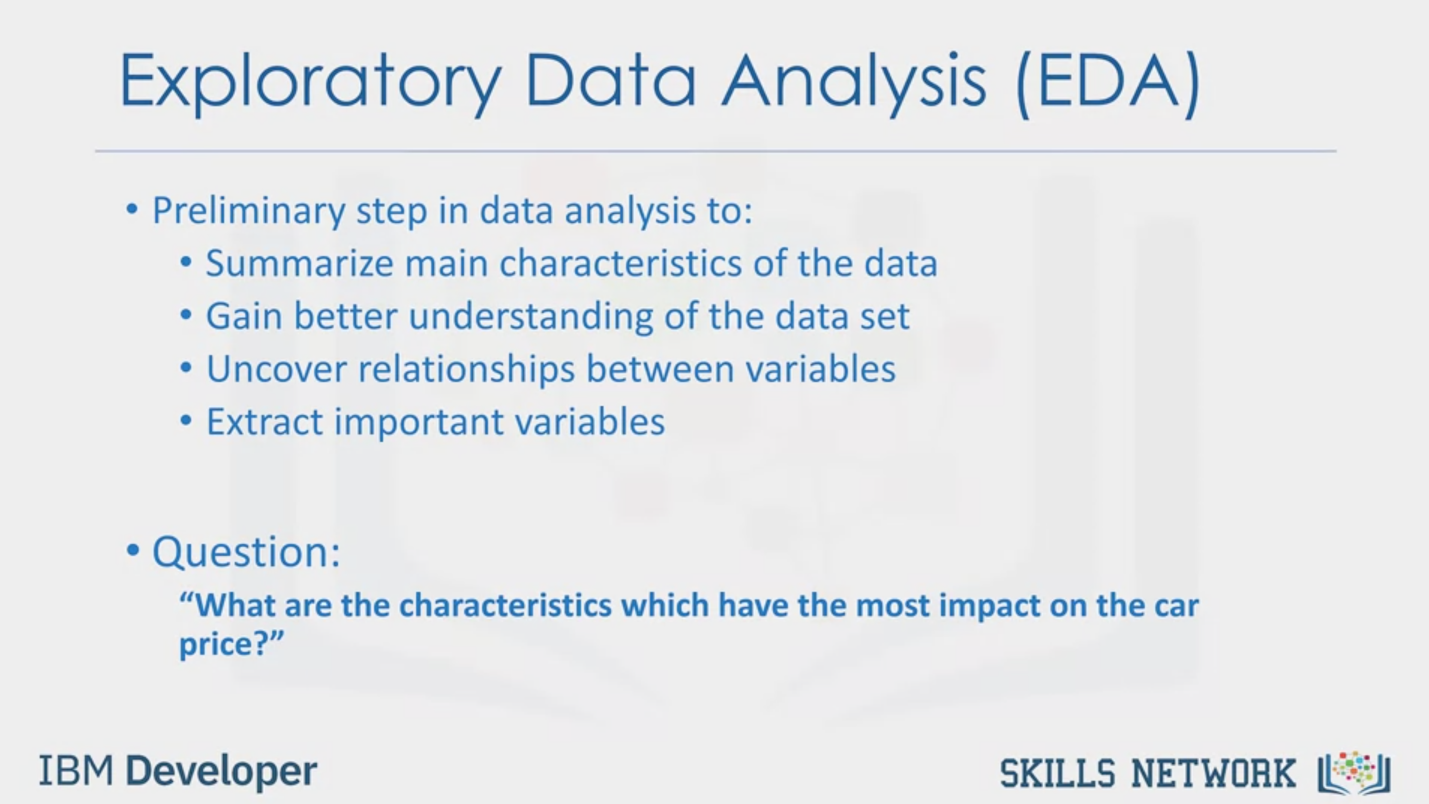
Exploratory Data Analysis

Module 3 Introduction

© IBM Corporation. All rights reserved.

IBM Developer

SKILLS NETWORK



Exploratory Data Analysis (EDA)

• Preliminary step in data analysis to:

• Summarize main characteristics of the data

• Gain better understanding of the data set

• Uncover relationships between variables

• Extract important variables

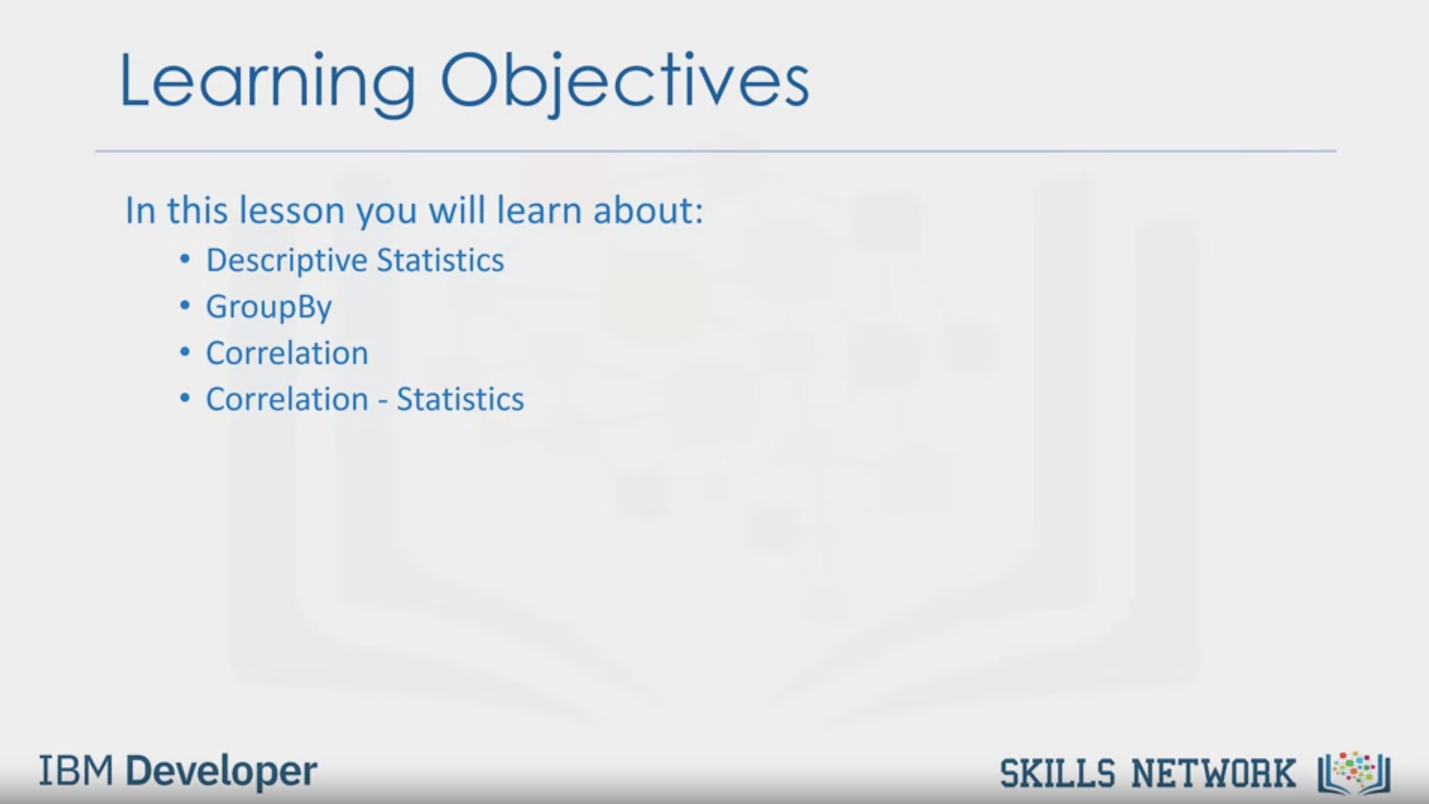
• Question:

"What are the characteristics which have the most impact on the car

price?"

IBM Developer

SKILLS NETWORK



Learning Objectives

In this lesson you will learn about:

• Descriptive Statistics

• GroupBy

• Correlation

• Correlation - Statistics

IBM Developer

SKILLS NETWORK



SKILLS NETWORK

IBM Developer

In this module we're going to cover the basics of exploratory data analysis using python.

Exploratory data analysis or in short EDA is an approach to analyze data in order to

summarize main characteristics of the data gain better understanding of the data set,

uncover relationships between different variables, and extract important variables for the problem

we're trying to solve. The main question we are trying to answer in this module is

what are the characteristics that have the most impact on the car price? We will be going through

a couple of different useful exploratory data analysis techniques in order to answer this

question: In this module you will learn about descriptive statistics, which describe basic

features of a data set and obtains a short summary about the sample and measures of the data. Basic

of grouping data using group by and how this can help to transform our data set, the correlation

between different variables, and lastly advanced correlation. Where we'll introduce you to

various correlation statistical methods namely pearson correlation and correlation heat maps