ubiqum code academy

Module 1 Data Analytics with Python: Understanding Customers

Data Analytics and Machine Learning

Storyline

- In this course students will be working for Blackwell Electronics as data analysts. The students' job is to use data mining and machine-learning techniques to investigate the patterns in Blackwell's sales data and provide insight into customer buying trends and preferences. The inferences students draw from the patterns in the data will help the business make data-driven decisions about sales and marketing activities.

- Students will use Python and various Data Analytics libraries to understand the relationship between customer demographics and purchasing behavior. Students will present their findings to management, explaining their insights and suggestions for data mining process improvements.

- The main goal of this unit is how to understand a dataset and learn how to handle one, this passes through an introduction to Python, to Jupyter notebooks, to EDA (Exploratory Data Analysis) and to data visualization.

Tasks

Task 1: Investigate customer buying patterns

Task 2: Investigate customer demographics

Data set

Through the whole program, you will be working with different datasets. You have to become familiar from the very beginning with how to manage a dataset. This is the first fundamental skill of a data scientist. A dataset is a flat-file with rows and columns. Each row is an individual element of the data.

The dataset is composed by transactions made either on-line or in-store over 4 regions. There is information also about the amount spent, the number of items purchased and the age of the customer.

Out[2]:

	in-store	age	items	amount	region
0	0	37	4	281.03	2
1	0	35	2	219.51	2
2	1	45	3	1525.70	4
3	1	46	3	715.25	3
4	1	33	4	1937.50	1

Each column is an attribute:

In-store
This attribute indicates if the purchase is made in-store or online

- 1 = true → in-store purchase
- 0 = false → online purchase

Age
The attribute indicate the age of the customer

Items
number of items bought in the same transaction

Amount
Amount spent in the transaction

Region
this indicates in which of the four regions the purchase was made

- 1 = North
- 2 = South
- 3 = East
- 4 = West

TASK 1

Data Analytics and Machine Learning

Module 1 – Task 1

Task 1: Investigate customer buying patterns

Message

From: Danielle Sherman

Investigating customer buying patterns

Hello,

As CTO and head of Blackwell's eCommerce Team, I'd like to welcome you aboard. I'm excited to get started on this project, but I'd first like to give you a bit of background to get you up to speed. Blackwell has been a successful electronics retailer for over three decades, with over numerous stores in various locations. A little over a year ago we launched our eCommerce website. We are starting to build up customer transaction data from the site and we want to leverage this data to inform our decisions about site-related activities, like online marketing, enhancements to the site and so on, in order to continue to maximize the amount of revenue we generate from eCommerce sales.

To that end, I would like you to explore the customer transaction data we have collected from recent online and in-store sales and see if you can infer any insights about customer purchasing behavior. Specifically, I am interested in the following:

- Do customers in different regions spend more per transaction? Which regions spend the most/least?
- Is there a relationship between number of items purchased and amount spent?

To investigate this, I'd like you to use data mining methods to explore the data, look for patterns in the data and draw conclusions. I have attached a data file of customer transactions; it includes some information about the customer who made the transaction, as well as the amount of the transaction, and how many items were purchased. Once you have completed your analysis, please create a brief report of your findings and conclusions and an explanation of how you arrived at those conclusions so I can discuss them with Martin.

Business questions:

- Do customers in different regions spend more per transaction?
- Which regions spend the most/least?
- Is there a relationship between number of items purchased and amount spent?
- Any other insights about customer purchasing behaviour

Deliverables:

A well commented jupyter notebook (.iypnb) so the mentor can review it

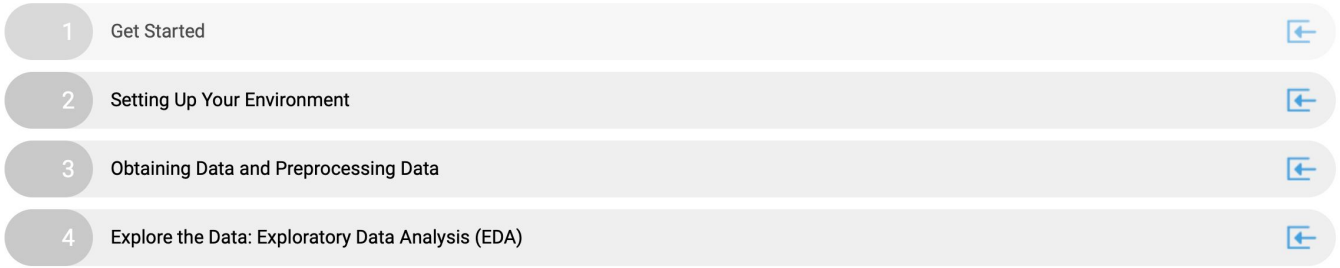
Deadline:

09.07.2021

Module 1 – Task 1

Task 1: Investigate customer buying patterns

Plan of Attack



1

What is the data science process

2

Install *Python* & Anaconda Navigator

3

Get the data, manage missing values, duplicates, outliers…

4

EDA – Explore Data Analysis. What are this data about?

Helpful resources

Type of Resource	URL
Installing Anaconda	Link
Example of EDA	Link
EDA Readings	Link Link
Correlation Reading	Link Link
Correlation Example	Link