Data Science Homework - Homework assignment for Data Scientist candidate

Objective: Develop a predictive model based on the provided Order and Online customer behavior data (data.zip). The analysis can be done in R or Python and should be presented in an R-Studio Notebook or Jupyter Notebook. The assignment should produce a multi-class classification supervised learning model to predict product category (prodcat1) a customer is likely to order. Use your expertise to design the analysis and provide a rationale of chosen approach. Once completed, please upload your assigment to your personal github repo and share the link. In your workflow, please touch on each of the following areas:

1. Exploration and understanding of the data sets
2. Feature engineering
3. Feature selection
4. Model design and sampling
5. Model generation
6. Model evaluation
7. Summary of results: 2-3 paragraphs textual summary

Note: It is not necessary to produce a highly predictive model, but, rather, to illustrate your understanding and practical knowledge of the model building process. There is no right answer, so you can go with certain number of assumptions about the data as you see fit. However, in case you're unable to proceed without the needed clarification, please feel free to reach out. Bonus: If you can work in customer segmenation as part of your EDA

Data Sets

Table order.csv 263278 obs. of 6 variables:

Columns	Data	Column Description
custno	int 18944 18944 18944 36096 1 6401 25601 57601 2 2 ...	Customer number
ordno	int 64694 28906 114405 62681 1 8187 41198 112311 70848 2 ...	Order number
orderdate	POSIXct, format: "2016-11-27 20:57:20" "2017-04-23 21:31:03"	Order date
prodcat2	int NA NA NA NA NA NA NA NA NA NA ...	Product category - detail
prodcat1	int 1 1 1 1 1 1 1 1 1 1 ...	Product category
revenue	num 76.4 130.7 139.2 72.5 100.2 ...	Revenue

Table: online.csv 954774 obs. of 7 variables:

Columns	Data	Column Description
session	int 419542 3030130 2638740 880408 2612179 880953 418956...	online session key
visitor	int 140970 14501 419353 90673 191542 419268 14938 419163...	Online visitor key
dt	POSIXct, format: "2016-09-16 05:03:23" ...	Online activity date
custno	int 3840 70400 21248 39168 47616 47616 47872 49920 49920 54784 ...	Customer number
category	int 1 1 1 1 1 1 1 1 1 1 ...	Online browsing category (prodcat1 from order.csv)
event1	int NA NA NA NA NA NA NA NA NA NA ...	Online event 1
event2	int 1 1 1 1 1 1 1 1 1 1 ...	Online event 2