← Go Back to Marketing & Retail Analytics

Course Content

MRA Project FAQs [DO NOT RELEASE TO LEARNERS]

For the MRA project milestone 1, can we use python, Tableau, and KNIME or do we have to use only one software?
Since still not much is taught regarding EDA through KNIME, that's why I asked whether we can use python as well.

You can use a combination of Tools -> KNIME, Python, and Tableau, or maybe all three together. No restriction on that.

Milestone 1 is based on Week 1 content -> Only EDA and RFM is expected

Milestone 2 is based on Week 2 content -> Only EDA and MBA is expected

Milestone 1

Could you please explain the significance of Days since the last order column?

You can compute Recency as "[Max(order date) - order date)]" as if you compare it with the current date, the Recency will have large numbers. so please use Max Order Date and that may be "30-05-2020" or "01-06-2020".

You can ignore the column "Days Since last order" and create yourself as stated above.
The RFM analysis needs to be done on Products or Customers?
Please perform RFM analysis using the Customer Name Field

- *Minimum 14 Slides of PPT required:*
  - *Minimum 8 slides on EDA and Inferences.*
  - *Minimum 2 Slides on RFM analysis*
  - *Minimum 4 slides on RFM inferences and identified segments*
  - *Do not add code*
  - *If you are using KNIME, adding KNIME Workflow is a must*
  - *Talk about Python package used in case you use python*
  - *It is a must to add plots/graphs in the PPT itself*

Should I preprocess the data in python first and then use that excel file in Tableau and KNIME or do I have to use the original excel file for all three.

You can preprocess the data first and then export the updated one to Tableau or KNIME.

Executive Summary will comprise of the combined insights from EDA and RFM or separately we have to

summarize for EDA and RFM.

You can summarise the EDA insights after EDA and RFM points after RFM. For Executive Summary, you can skip it or mention a few top points (4 lines from each EDA & RFM). It has been made non-mandatory for evaluation.
In question 3 Customer Segmentation using RFM analysis (4 segments) here 4 segments are asked, please explain the segmentation means Bins which we are making in RFM
4 Segments means 4 Bins (1,2,3,4) like Recency - 4 being most recent or for Monetary or Frequency, 4 being high monetary value or most frequent.

Please help to understand the flow. As per my understanding, we first have to do summary statistics and EDA using python and DVT. In DVT do we have to make dashboards and storyboards or just graphs and paste them in PPT?

You have to perform EDA using either python or Tableau or Both.
You have to put the plots in the PPT.
No need for Story/Dashboard. But do put the Tableau Public link of your Tableau workbook in case the evaluator wants to refer to it (Rare).
Just create meaningful and visually appealing graphs using Tableau or Python or a mix of both for the EDA part.

In question 3 it says mention which tool have we used. Would want to know what all tools are there.

> Customer Segmentation using RFM analysis (4 segments) -> Which tool used? -> What all parameters used and assumptions made? ->If KNIME used, Workflow image to be put -> Output table head

Which tool used means which software (KNIME or Tableau or language Python) is used to get the RFM Analysis done.
i.e. If KNIME is used, you should write " KNIME is used to perform the RFM Analysis and here is the workflow diagram" [put workflow screenshot]. Discuss the support & confidence values threshold put (for Milestone 2). Write about the Recency, Frequency, and Monetary calculation formula used [for Monetary -> sum(sold_products) customerwise]. Rest you can guess as we cannot go in deep for an ongoing project. Put the head of the output of the RFM analysis Table.

If KNIME is used then we have to show workflow. Does this mean those yellow connected boxes? But what is the purpose of this in PPT?

Yes, it is important to see what all nodes were used in the workflow to get the output. We can see if the learner has followed the steps we showed in the videos. If you use python, please put the python package name used to get RFM done. You can write " I used python and employed so & so package[take the definition from documentation] to get the answers and here are the threshold values used.

For question 4, do we take the output in Excel and show the data in tableau and paste those graphs in PPT regarding best, loyal customers, etc?

You can take the RFM output in Excel and then analyze it in Excel only, or Tableau. That's up to you. Or you can simply use Excel to sort the customers and put the head of the sorted (as needed) table in PPT. While explaining the criterion why these are the best, or loyal or...?

So overall confused about steps and what should the PPT look like. Please help to understand the flow as we do not have a sample PPT.

Attaching a skeleton of PPT. You can refer to it. It will give you a starting point. You can customize the theme and use better fonts. Make sure all your graphs are visible and readable. Often tableau graphs have greyish labels which are not visible, the evaluator may deduct marks if not readable. Please take care of aesthetics as well. This is created just to give you a starting point.

Milestone 2

EDA Part for Milestone 2 is similar to Milestone 1

- *Please do write about Market Basket Analysis (1 slide is enough). What to write is something you have to decide (Definitions and usage, parameters used)*
- *Please mention the thresholds you have taken for deciding the support and confidence*
- *Please put the Association Rule output (about 15-20 rows) in the PPT*
- *Please discuss the values of support, confidence, and lift of the Association Rules Output and interpret the rules generated*
- *Please put about 3-4 Recommendations and 3-4 Example Discounts that can be given to the customers*.

I am getting a blank table after running Association Rules. What should I do?

Please try to adjust support & confidence values in the Association Rule Miner Node Config. Do not go too low while setting these values as it may give **a "Java Heap" Error**. If it gives an error at some value, then you know what value not to keep.

Ex: support = 0.05 and conf =0.50 in this dataset may give results but may not for other datasets.

Generic

You will be evaluated based on the PPT and aesthetics of the PPT as well. Please be professional in making the presentation. Loosely made PPTs will be penalized. Please put all the charts/inferences/ outputs in a readable format in the PPT only.

Can I work on Tableau for Graphs and KNIME for MBA & RFM?

Yes, this project is tool independent and you can use any tool covered in the Course like Python or R or KNIME, or Tableau.

Can I clean the dataset?

Yes, you can pre-process the dataset.

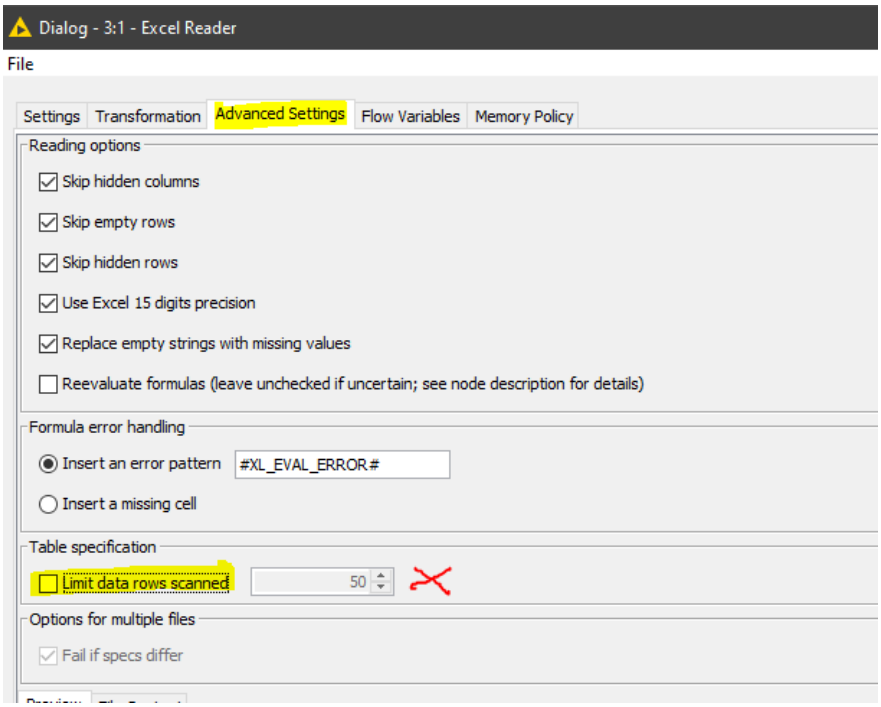
What do I need to upload?

You need to create a PowerPoint Presentation for analysis performed by you. You need to put the output graphs, tables, and write inferences drawn from the outputs. You also need to upload the code file of the respective tool used.

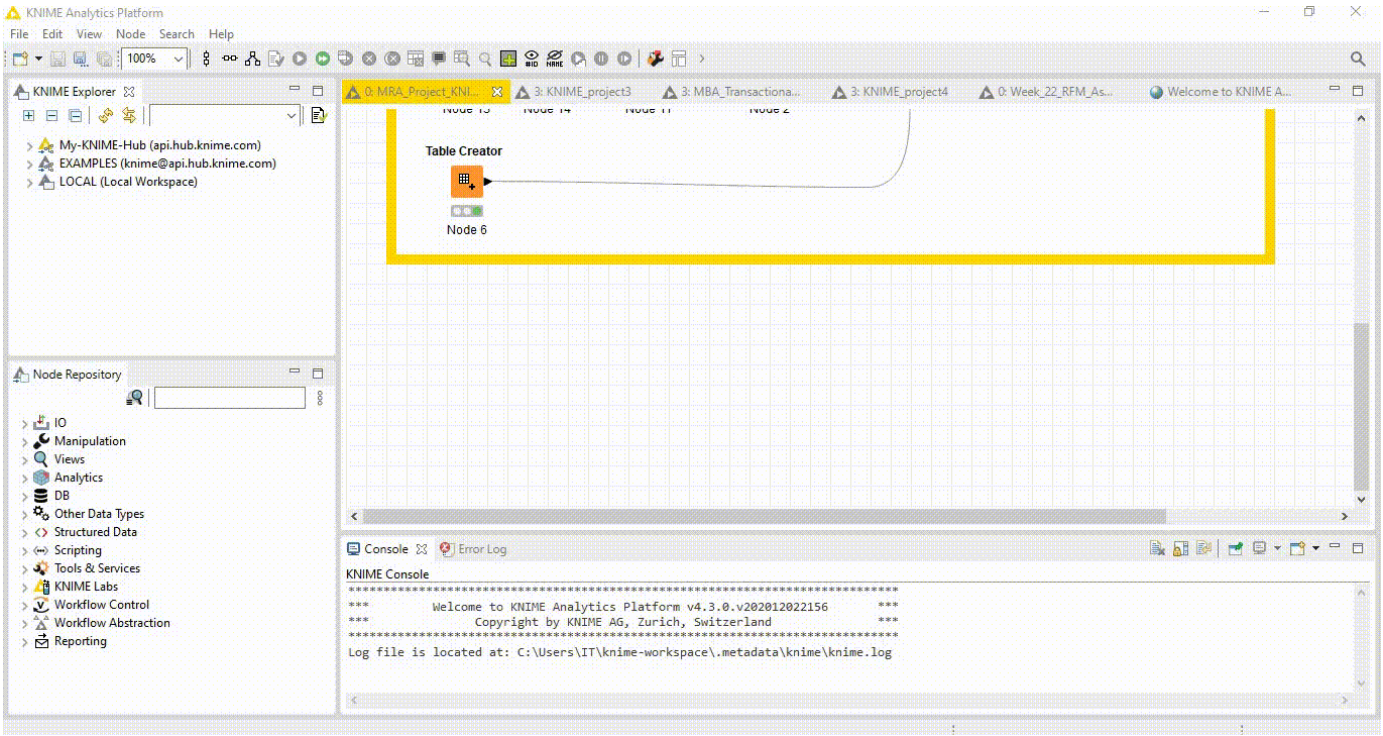
For example, you created plots in a tableau then you need to put the image of the plots in the PPT and explain them. A similar approach for other tools.

In the case of Tableau Public, please put the tableau public link in a separate document as you did in the DVT Course project.

KNIME Version 4.3 or above will give an error while reading the dataset. Please uncheck the following in the excel reader node's advanced settings:



Exporting KNIME Workflow for sharing with us while submission:



Previous