Branch: master ▾ Find file Copy path

learner / Exercise02.md

cnukaus Update Exercise02.md

500df60 a minute ago

1 contributor

Raw Blame History

41 lines (21 sloc) 1.6 KB

Pre-reading

- 3 min read https://www.dimodelo.com/subscribers/Data_Warehouse_Bus_Matrix.zip
- 5 min product evaluation: https://datastudio.google.com

Homework

Do data preparation (same methods as we did in Lecture 02), which will

..* fill all blank values (as our Python notebook) ..* Create a metadata description table (as ourr df_metadata) ..* Obtain results for each continent (note this definition is missing, you have to define it in your own reference table), so we have results like: Each decade, each continent, the average total injures ..* a stacked bar chart, to show monthly count of injuries by each maker (use color legend to tell a make)

in a new tool such as

..* Google Data Studio, Power BI or Tableau (note it has a separate data transform tool)

..* R, Spark

..* Alteryx, Knime, Orange or WEKA

Then save to local Exercise02 folder and commit, pull request

Bonus 作业

create a Slack or Discord bot and push your Homework results to a channel

Reference

翻转课堂的可汗学院

CRISP-DM方法论再认识

"分析规划宜采用"业务导向+数据驱动"的方式（如下图所示）。从关键业务目标分解出发，关联到具体的业务领域（研发、建设、运行、运维、安全环保、销售、采购等），从重要度和紧迫度的角度，对可能的业务分析问题进行评估。然后，结合初步的因子分解，评估每个题目的所需数据的完备度（Readiness）。综合业务价值和数据完备度，进行多个项目的优先排序"