Post datasets that you found interesting here and propose a business problem we could aid in solving with analyzing said data.

Fuzel: Generic sales data from previous project

Lauri: We could mix datasets such as Tableau Superstore Data (with modifications to fit the agenda, such as naming products to renovation or design products and services) and for example U.S. housing sales data to aim for a market operation optimization (where to focus sales and marketing efforts)

For generic sales stuff these seem like good datasets:

KPMG Virtual Internship <https://www.kaggle.com/datasets/neharautela/kpmg-virtual-internship>

Sales Domain <https://www.kaggle.com/datasets/ad043santhoshs/sales-domain?select=Sales_domain.csv>

Flipkart product dataset: <https://www.kaggle.com/datasets/priyankkhanna/flipkart-product-dataset-by-priyank-khanna>

Ideas for a business problem:

Sales forecasting, Customer segmentation, Operational efficiency,Financial analysis, Inventory management, Fraud detection, Risk management, Website performance, Supply chain optimization, Employee performance, Social media analysis, Customer churn.

Juho: I didn’t find Tableau data very interesting, but I suppose if we can think of an interesting problem then that wouldn’t be as important. I looked at some of the Kaggle competitions in the past to get some ideas, one that looked promising was time series forecasting problem found here: <https://www.kaggle.com/competitions/store-sales-time-series-forecasting/overview>

ML solutions might be a bit out of scope for this course, but there’s still lots of sales data in there, along with related and complementary datasets such as holiday events and oil prices (economically impactful) that could be used for visualization and analysis.

Lauri: Since there wasn’t really any other ideas I would go with the set that Juho found on Kaggle. I propose that we could work on trying to analyze the situation of sales of Favorita stores in Ecuador. There is no need for ML solution, visual analysis is enough and I think this is a good place to start.

Elmeri: Nothing new really to add on top of the ideas previously listed here and what we discussed during the meeting. As mentioned above, with the scope of this course, I think a simpler analysis is more than enough. In the Project Description this was mentioned “...find datasets that complement each other for a more comprehensive project (highly recommended)”, so we’ll have to think whether that’s possible or not. I also think the dataset Juho brought up seemed interesting and could be implemented in the project. I think there’s plenty to choose from though, some interesting topics in the Unicef and WHO datasets as well.