E05 – Visualization (Metabase)

Business Intelligence

Exercise

Winter Term 2025/2026

Agenda

- Introduction
  - Introduction
  - Notes on Software
- Exercise
  - Tasks
- Credits and Materials

Introduction

The main goal of the “Visualization” exercise is to introduce students in the world of the business intelligence and data analytics using one of the available out there data analytics tools. Usually, various kinds of visualization tools are used to deliver to decision makers the results of an analysis in form of reports, dashboards, ah-hoc queries. Despite this set of exercises focused on a limited number of software tools, obtained knowledge and hands-on experience could be very useful and could be easily re-used for other similar software solutions.

In this exercise you are going to use “Metabase” software.

Notes on Software

See the materials from the previous exercise **E02, E03**

Additional instructions on the infrastructure (online platform) for this course can be found in
E00 - Infrastructure

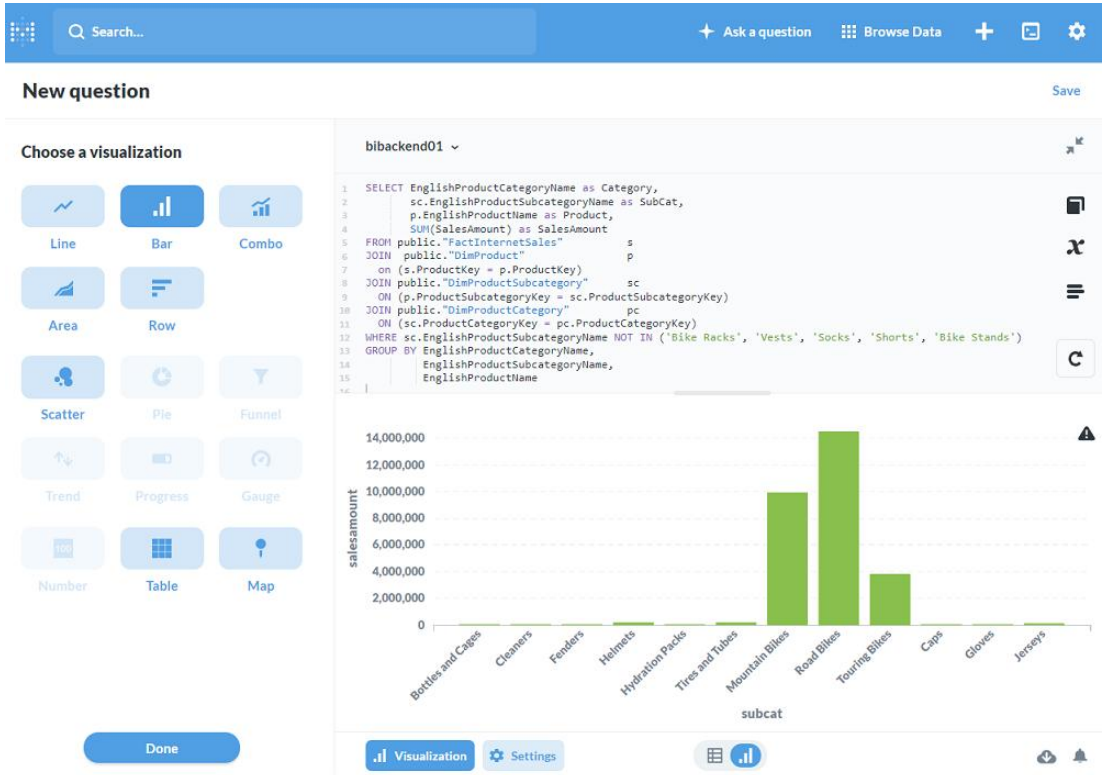
Start

- In this exercise the “Metabase” visualization tool is used
- There are 3 possibilities/options to get access to this tool
  - (option 1) use provide to you instance of Metabase - for further details refer to “**E00 – Infrastructure.pdf”**
  - (option 2) use docker - https://www.metabase.com/start/oss/
  - (option 3) use jar file - https://www.metabase.com/start/oss/
- You will need a DBMS to work with this data
  - You could use offered infrastructure
    - For further details refer to “**E00 – Infrastructure.pdf”**
  - Note that the data are already available, if not that
    - (option 1) import the data into your database using web UI
    - ~~(option 2) see the schema **bicourse_db_demo**~~
    - (option 3) import the data use SQL statements or CSV files into your own database
- About data
  - To import data use SQL statements, CSV files or Python script provide to you
  - The table fields and data type should be self-descriptive

Exercise

Description - Metabase

Metabase is a simple and powerful online analytics tool that lets anyone discover insights and make decisions based on their and external data. It easy to use tool with the number of important features, which facilitates smooth visualization experience and story telling with data



Technical Hint: Start your metabase

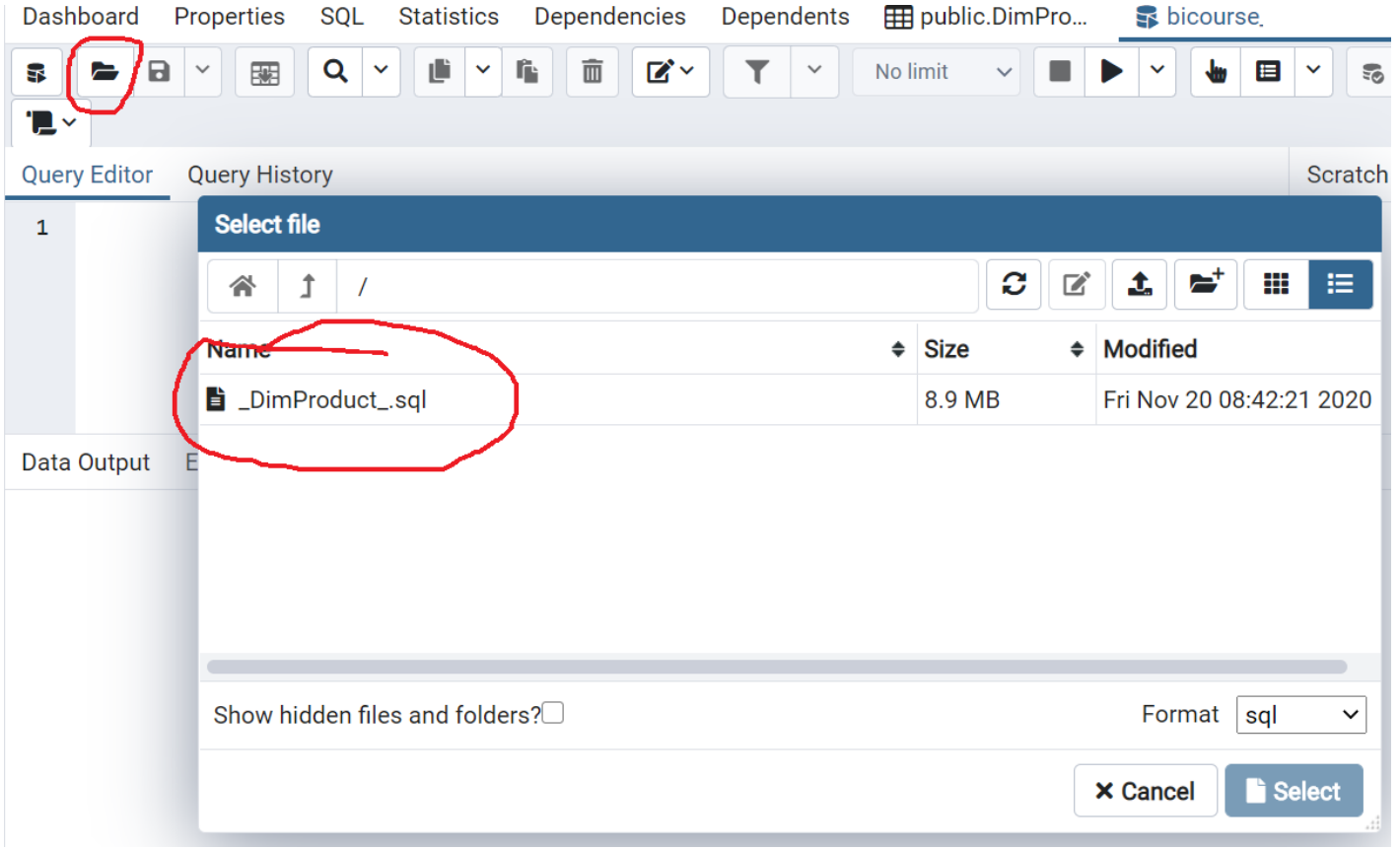
Don’t forget to start you metabase instance within the provided infrastructure before you start to work with it.

(optional) Technical Hint: Import Dataset Using web UI

One could use a web UI to import datasets directly into own database schema for further usage.

(optional) Technical Hint: PGAdmin4 and SQL File to Import

It is not always possible to move (e.g. „copy/paste“) a lot of data at once between your clipboards and PGAdmin4 SQL tool. Thus, before you will import the data from a big SQL file with inserts, upload this file first to the server. Usually, you should be able to upload files up to 50 Mb.



(optional) Technical Hint: Upload CSV Data into PostgreSQL using Python

If you are struggling to upload the data into PostgreSQL, you can use following Python script (you will need to install pandas package and Python). *You should also know, that copying following code into your Python environment may now work directly, because Python's uses tabs/spaces for it's formatting.*

```
# pip install pandas
# pip install sqlalchemy

import pandas as pd

DB_CONNECT_URI = 'postgresql://USER:PASSWORD@DB_URL/DB_NAME'

def import_into_table(csv_fname, table_name):

    print ('[i] Importing following file into database: {0}'.format(csv_fname))
    df = pd.read_csv(csv_fname)
    df.columns = [c.lower() for c in df.columns] #postgres doesn't like capitals or spaces
    print (df.columns)

    from sqlalchemy import create_engine
    engine = create_engine(DB_CONNECT_URI)

    df.to_sql(table_name, engine)

import_into_table(csv_fname='data/listings.csv', table_name='airbnb_listings_summary_01')
```

NOTE: *You can also use provide to you database schema (demo) "**bicourse_db_demo**" to access certain data. Be aware the you are only able to access the data in the "read-only" mode*

Tasks

Tasks – Part 0

1. Get access to a metabase instance. There are 3 possibilities to get access to this tool / instance

   - (option 1) Use provide to you instance of Metabase - for further details refer to “**E00 – Infrastructure.pdf”**
   - (option 2) Use docker - https://www.metabase.com/start/oss/
   - (option 3) Use jar file - https://www.metabase.com/start/oss/

Tasks – Part I

0. Get yourself familiar with the information provided by the original “Getting Started” tutorials and documentation from Metabase and also other sources:

   a) https://www.metabase.com/learn/metabase-basics/getting-started/index
   b) https://www.metabase.com/docs/latest/dashboards/introduction
   c) Video tutorials
      - (single) Metabase tutorial: everything you need to know to get started in under 40 min
        - https://www.youtube.com/watch?v=LJ0l2HZ8Lp8
      - (serial) Tutorials: Getting started with Metabase
        - https://www.youtube.com/playlist?list=PLzmftu0Z5MYGY0aA3rgIGwSCifECMeuG6

1. Please, answer following question - what is "automatic explorations" in the terms of Metabase?
2. Please, answer following question - what means “asking a new question" in the terms of Metabase? How can you utilize this functionality of this tool?
3. Please, answer following question – what types of the visualizations this tool provides?
4. (optional) Create a simple dashboard as described in the tutorial (you may another dataset as well)

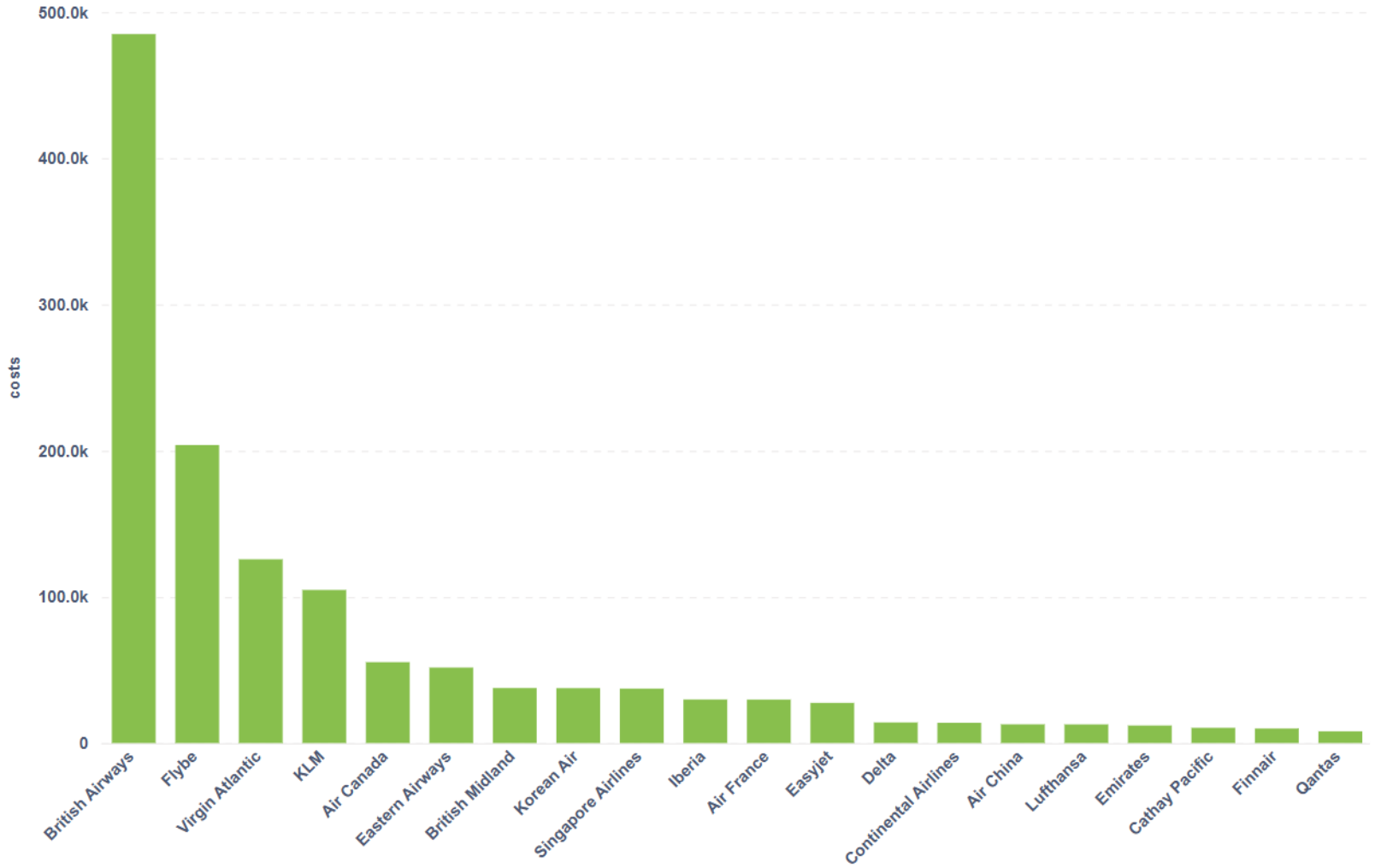
Tasks – Part II

1. Import “flights” dataset into the database
   - (option 1) use “auto import” feature offered by the course infrastructure
   - (option 2) download and import manually
     - “flights” dataset is available here
       - https://raw.githubusercontent.com/apache-superset/examples-data/master/tutorial_flights.csv

2. What kind of data / fields the dataset contains? List them (name + data type)

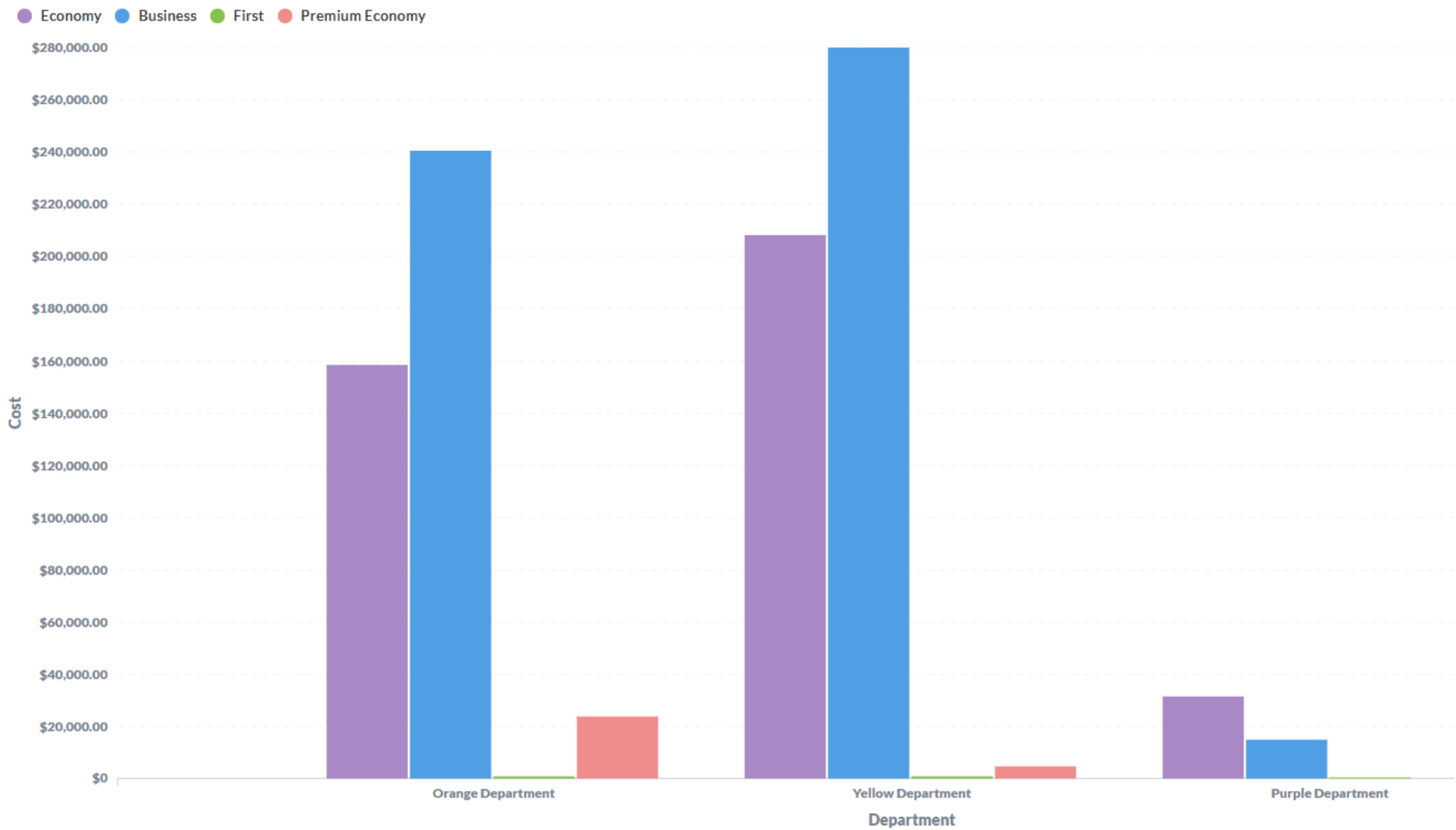
Tasks – Part II

3. Reproduce this plot by using “red” color for the bar chart and limiting airlines to 20



Tasks – Part II

4. Reproduce the given plot. Which insights you can spot?



Tasks – Part II

5. How many countries are represented in the dataset? Visualize the number and countries
6. Compare and visualized the amount of sold tickets (single / return).
7. Show, how the given costs are distributed over the time.
8. Create a geo-map with the original latitude / longitude.
9. Show on the geo-map the farthest destination from “London Heathrow Airport”. Display both locations on the map.
10. Create a new dashboard that contains following information
    - Total amount of flights available
    - Single / Return tickets ration from June and July
    - Location of the busiest airport
    - Costs for each department

Submission

- Use StudIP to upload your solution (PDF report)

- You should upload one of the following document
  - Report as a PDF file with your solutions
  - Each answer (e.g., SQL query) should be annotated by you
    - Original question (+ number of the task)
    - Answer

- Name convention for your submission file (without extension)
  - **E05**_FIRSTNAME LASTNAME

- Submission deadline (it is a “soft” deadline)
  - 10 days after this exercise starts
  - Some exercises could take a bit more time and could be submitted later
  - NOTE: to receive feedback, you should first submit your progress