

Business Intelligence and Applications

Assignment 2: Pivot Tables & Tableau

Dataset

The `us_cities_dataset.csv` file, which you can download from the iCorsi webpage for the course, contains information about cities in the United States. The data is sourced from the U.S. Census Bureau, U.S. Bureau of Economic Affairs and Kaggle.

The following list is a short overview of the variables present in the dataset.

- **city**: name of the city
- **state_id**: unique identifier for the state the city is located in.
- **state_name**: name of the state the city is located in (it should map with `state_id`).
- **county_name**: name of the county (administrative subdivision in-between city and state) the city is located in.
- **lat**: latitude of the city.
- **lng**: longitude of the city.
- **population**: population estimate.
- **density**: population density.
- **military**: boolean value that identifies if the city is under military administration.
- **incorporated**: boolean value that identifies if the city is incorporated, i.e., if it is allowed to have elected officials, or alternatively is run directly by officials nominated by the State.
- **ranking**: between 1 to 5, identified the type of city . It can be thought as an identifier of the type of municipality: metropolis (1), mid-size city (2), town (3), village (4) and community (5).
- **available_land**: square feet of available land area in the city.
- **available_water**: square feet of available water area in the city.

- **mean_income**: average income of residents in the city, in USD.
- **median_income**: median income of residents in the city, in USD.
- **stdev_income**: standard deviation for the income of residents in the city, in USD.
- **airport_count**: number of airfields, i.e., airports and heliports, in the city. This considers all legal places where an air vehicle can take off from, e.g., local airport just like international airport.
- **average_airport_score**: average for the “airport score” in the city, i.e., a score identifying how much air traffic a certain airport has. This is averaged over all airports for a city.
- **average_airport_elevation_ft**: average elevation of airports inside the city limits.

Task

The task for this assignment is to use the tools shown in class (Excel Pivot Tables and Tableau) to analyse and summarise the dataset and to apply roll-up and drill-down operations, to gain an insight into the data. You should also explore the possibilities and tools provided by Tableau and compare them to what is achievable with Excel Pivot Table. In particular, you have to include (but not limited to) the following points in your analysis:

1. Analyse the data and identify the state with the most cities in it, and that with the most airports.
2. Find counties with the highest population in them: are they the counties with the most cities in them as well? Compare the findings. For these counties, show also the distribution of the rankings.
3. Which are the fairest cities, i.e., cities where the median and average income are the closest, to live in? You can compute a “fairness score” as the difference between the average income and the median income. What can you say about the fairness score and the mean income, e.g., the richer the population of the city, the less fair it is? Draw your own conclusions. Describe also the relationship between the fairness score and the city’s population.
4. Geographically visualize the cities and their population on a map. If plotting all of the cities becomes too computationally expensive, you may consider to cut only up to a certain number, e.g., top 100 cities (by population) in the US.
5. Is there any correlation between (1) the population and the available land; (2) median income and average airport score and (3) the available water area with the average elevation of the airports in the city?

6. Perform analysis at the state level: which are the most populous states? How many counties and cities do they have? Visualize these findings on a map as well.¹
7. Can you observe any particular trends in the data? Are there any significant insights that you gained while analysing the dataset?
8. Compare the usage of Excel's pivot tables with the features offered by Tableau. Which tool is more suitable and convenient for which types of tasks? Is there anything that was possible to do with Tableau that you couldn't do with pivot tables and the other way round?

To install Tableau, please follow the link on the iCorsi page. For Microsoft Office, you can download it through the University (see the official guide).

Submission Rules

You can work on this assignment in a group of maximum 4 people. If you successfully passed "Pivot Tables and Tableau" assignment last year, there is no need to redo it²

Prepare a PDF report of up to 3 pages, including tables (or excerpts thereof) and graphs, succinctly describing the salient aspects of the dataset and your findings. You should specify your names on the first page of the report.

Please, also prepare and record a short (max 6 minutes) video presentation where you describe your solution, approach, and findings. In your presentation, you should guide us through the steps you took to acquire the results and build the charts, while focusing on important details and verbally explaining the proceedings. You can use Microsoft Teams / Zoom (or any other tool of your choice) to arrange a meeting and record your presentation. Or you can use your own tools of choice.

You should upload the recorded presentation to any cloud storage (Dropbox, Google Drive, etc.) and include the link to the file in your PDF. This should be done only by a single member of each team. Please, indicate your names in the uploaded file's name.

The report must be submitted by one of the members before Monday, 03 April 2023, 2:15 PM (at the latest) through the corresponding submission point on iCorsi platform.

You will fully pass the assignment if you satisfactorily solve and present at least 6 tasks; you get a partial pass if you solve at least 4 tasks. A selection of the submissions will be invited to present their solutions live to the class during the lecture held on April 24, 2023.

Should you have any questions regarding the assignment or problems with your submission, please contact the TA (leonardo.alchieri@usi.ch).

¹Tableau should allow you to draw automatically geographical maps based on the state name.

²Contact the TA if this is the case.