



Decision Theory and Analysis MS5104

Lecturer: Dr. Anastasia Griva

- Group assignment -

Assignment scope

Using Tableau, the objective of this assignment is to design a data visualisation dashboard, implementing best practice in data visualisation design. As well, your critical thinking, decision making and storytelling capabilities will be assessed.

Assignment description

Imagine that the C-level executives (CEO, CPO, CIO, CFO etc.) of a fictitious company asked you to investigate data and share your experience with a view to find a new opportunity (and/or pinpoint potential threads) for this business (e.g. entering a new market, developing a new service, attracting new customers, trying to avoid customer attrition etc.). They asked for your assistance in making such a decision and requested that you analyse a dataset using Tableau and present a compelling argument to invest in a specific market or innovation.

In this group assignment you will need to design a single dashboard visualisation for a dataset / open dataset of your choice, using the Tableau software (see next section for indicative open datasets).

Based on the selected data try to present evidence on whether the company can invest in any specific market or innovation, to assist the company in this decision. Formulate questions based on the given dataset to assist you in presenting a nice story using your data. Use your critical thinking and pay particular attention on storytelling. Your task is to design an interactive visualisation dashboard that you believe effectively communicates the data and present insights e.g. trends, opportunities within a business context.

You are free to make any assumptions e.g. about the fictitious company, its products, market etc.

Your dashboard should:

- Allow some level of interaction and data interrogation (i.e. include filters)
- Be designed in accordance to data visualisation best practice (please reference the (i) Tableau tutorial material and (ii) lecture content and material on Blackboard)

During your analysis pay attention to the choice of visualisation type, size, colour, scale, and other visual elements. Keep in mind that all these decisions facilitate effective communication to potentially support decision-making.

Indicative datasets

Below you may find some examples of links to online and open datasets:

- Ireland's Open Data Portal
Promoting innovation and transparency through the publication of Irish Public Sector data in open, free and reusable formats
<https://data.gov.ie/data>
- European Data Portal
Thousands of datasets from EU member states regarding issues of public interest such as Transport, Agriculture, Economics, Environment, etc.
<https://data.europa.eu/en>
- Kaggle
Includes over 50,000 public datasets
<https://www.kaggle.com/>
- United Nations Data Sets
Several databases containing over 60 million records covering a wide range of themes including Agriculture, Crime, Education, Employment, Energy, Environment, Health, HIV/AIDS, Human Development, Industry, Information and Communication Technology, National Accounts, Population, Refugees, Tourism, Trade, as well as the Millennium Development Goals indicators.
<http://data.un.org/Explorer.aspx>
- UK Government Data
Datasets available from all UK central government departments and a number of other public sector bodies and local authorities.
<https://data.gov.uk/search>
- Mendeley Data
It includes 28.9 million datasets from domain-specific and cross-domain repositories
<https://data.mendeley.com>
- Irish Organisation for Geographic Information (IRLOGI)
Irish GIS data sources and links to other portals
<http://www.irlogi.ie/datasets.php>
- Spatial Data & Other Datasets for Ireland (link maintained by UCD Library)
Links to hosts of datasets for such topics as:
Administrative Boundaries; Agriculture; Air Quality; Archaeology; Bathymetry; Birds; Built Heritage; Conservation Areas; Crime; Economics; Education; Elevation Data; Extractive Industries; Flooding; Flora; Forestry; Geology; Habitats; Health; Housing; Hydrology; Insects; Ireland_Shapefiles; Labour Market; Land Cover; Land Use; Mammals; Marine Data; Mining; Natural Heritage; Satellite Imagery; Population; Social Data; Soil; Topography; Transportation & Travel; Utilities; Waste; Water Quality; Weather & Climatology; Wind Energy.
<https://libguides.ucd.ie/gisguide/FindSpatialData>

- Central Statistics Office Databases (Ireland)
Links to statistical databases of the Irish CSO, other public sector bodies, and international organisations.
<https://www.cso.ie/en/databases/>
- EPA Geoportal site
This geoportal is designed to make data about the environment easier to find, browse and understand. You can download GIS data and find other portals to download datasets.
<https://gis.epa.ie>
- Hadoop Illuminated > Publicly Available Big Data Sets
List of publicly available big data sets
http://hadoopilluminated.com/hadoop_illuminated/Public_Bigdata_Sets.html
- KDnuggets
Links to several business data sets from North America, Europe, and worldwide databases <https://www.kdnuggets.com/datasets/index.html>

You are free to select any dataset of your choice, even if it is not included in the aforementioned suggestions. You are also free to select a smaller sample of the dataset under examination in case you face any issue with the processing capabilities of your PC/laptop.

Deliverable and deadline

Intermediate Deliverable

All teams are required to create a "Group wiki" within their group and denote/post the following:

- which dataset you selected to analyse for your assignment
- why (in 2-3 sentences)
- which are your business questions

This is compulsory for your assignment.

DEADLINE: 7/10 @16:00

Final Deliverable

To describe your findings, visualisation design, and recommendations to the C-level executives of the fictitious organisation, your group is required to submit a **15 minute presentation (max)** describing your analysis. No report or ppt is required.

You can either:

- present the stories you have built in Tableau
- or flip between ppt and Tableau
- or (less preferably) include Tableau screenshots in a ppt.

Do not forget that you present the results to business people some of them are having a little to no knowledge on analytics. Your presentation will aid them to support decision making.

Do not include technicalities in the main presentation.

The deliverable should be submitted through Blackboard.

DEADLINE: 18/11 @16:00

Marks awarded : 40% of the overall marks for the course.

One person for each team is responsible to submit the assignment.

All team members should present.

Peer evaluation forms should be submitted by all group members.

The first slide of the presentation should include: (i) group name, (ii) team number, (iii) student names, (iv) student ids.

A professional photo of the group members is recommended, but it is not compulsory.

Note: Plagiarism detection software will be used during the assessment of the submissions.

Late Submission or No Submission

Late submissions will be penalised by 10% per day (or part of). Submissions received later than 3 days past the original submission date will be regarded as a non-submission and will be awarded a zero mark. There are no facilities or opportunities to re-submit the assignment before the next offering of the module, should student(s) fail to submit by specified deadlines in the first instance, or find that students are sitting the written examination at Autumn.

Marks for continuous assessment (including zero marks for non- submission) will be carried forward to Autumn.

Groups formation

You will be assigned into groups of 4-5 (more rarely) members by the lecturer. Teams having 5 members should work more on the presentation.

You are required to name your group, and use this name to any queries you send to the lecturer. Be creative with the names!

A peer evaluation form should be completed by all group members.

Summary of requirements

- Download tableau (instructions have been provided on blackboard)
Complete the Tableau online tutorials
- Select any open dataset (organisational, environmental, government, etc.) of your choice. Explain rationale for the selected dataset and consider how it might be used for a fictitious organisation to demonstrate the opportunities or threats in investing in a specific market or innovation of your choice.
- Report on key data insights from your selected dataset. Your group must set the assignment questions (i.e. what is it that you wish to present with the dataset) and reach agreement on your interpretation of the dataset.
- Explore a few options for the open datasets. Identifying a suitable open dataset can prove to be the most challenging part of the assignment.
- You must place emphasis on generating rich insights on the selected dataset and 'tell the story' to address your research questions, e.g. generating business value insights on new opportunities/trends from your dataset.
- Pay attention to story-telling and visualisation.

Assessment criteria (40% of final mark)

The video presentations will be assessed on a number of factors including:

- Your explanation of the dataset (e.g. decision to select the dataset, context of dataset, preparing the dataset, etc.)
- Your ability to determine suitable questions based on a selected dataset (e.g. explain group decision to create a number of questions which your dataset can adequately address and why these questions are worth exploring)
- Your ability to demonstrate your knowledge of Tableau functions
(Note, not all of these functions are necessary, but you should select the most appropriate to support your insights).
- Your ability to create worksheet visualisations (a number of Tableau worksheets to explore research questions).
- Your ability to develop dashboard (experimenting with various visualisations to represent clear key data insights)
- Your ability to professionally present stories (explain the business value)
- Your presentation skills 15-minute group video presentation

An interesting article:

Harvard Business Review (HBR)

<https://hbr.org/2012/10/data-scientist-the-sexiest-job-of-the-21st-century/>

MODERN DATA SCIENTIST

Data Scientist, the sexiest job of 21st century requires a mixture of multidisciplinary skills ranging from an intersection of mathematics, statistics, computer science, communication and business. Finding a data scientist is hard. Finding people who understand who a data scientist is, is equally hard. So here is a little cheat sheet on who the modern data scientist really is.

MATH & STATISTICS

- ☆ Machine learning
- ☆ Statistical modeling
- ☆ Experiment design
- ☆ Bayesian inference
- ☆ Supervised learning: decision trees, random forests, logistic regression
- ☆ Unsupervised learning: clustering, dimensionality reduction
- ☆ Optimization: gradient descent and variants

PROGRAMMING & DATABASE

- ☆ Computer science fundamentals
- ☆ Scripting language e.g. Python
- ☆ Statistical computing package e.g. R
- ☆ Databases SQL and NoSQL
- ☆ Relational algebra
- ☆ Parallel databases and parallel query processing
- ☆ MapReduce concepts
- ☆ Hadoop and Hive/Pig
- ☆ Custom reducers
- ☆ Experience with xaaS like AWS

DOMAIN KNOWLEDGE & SOFT SKILLS

- ☆ Passionate about the business
- ☆ Curious about data
- ☆ Influence without authority
- ☆ Hacker mindset
- ☆ Problem solver
- ☆ Strategic, proactive, creative, innovative and collaborative

COMMUNICATION & VISUALIZATION

- ☆ Able to engage with senior management
- ☆ Story telling skills
- ☆ Translate data-driven insights into decisions and actions
- ☆ Visual art design
- ☆ R packages like ggplot or lattice
- ☆ Knowledge of any of visualization tools e.g. Flare, D3.js, Tableau

