

Descriptive Analytics & Structured reports

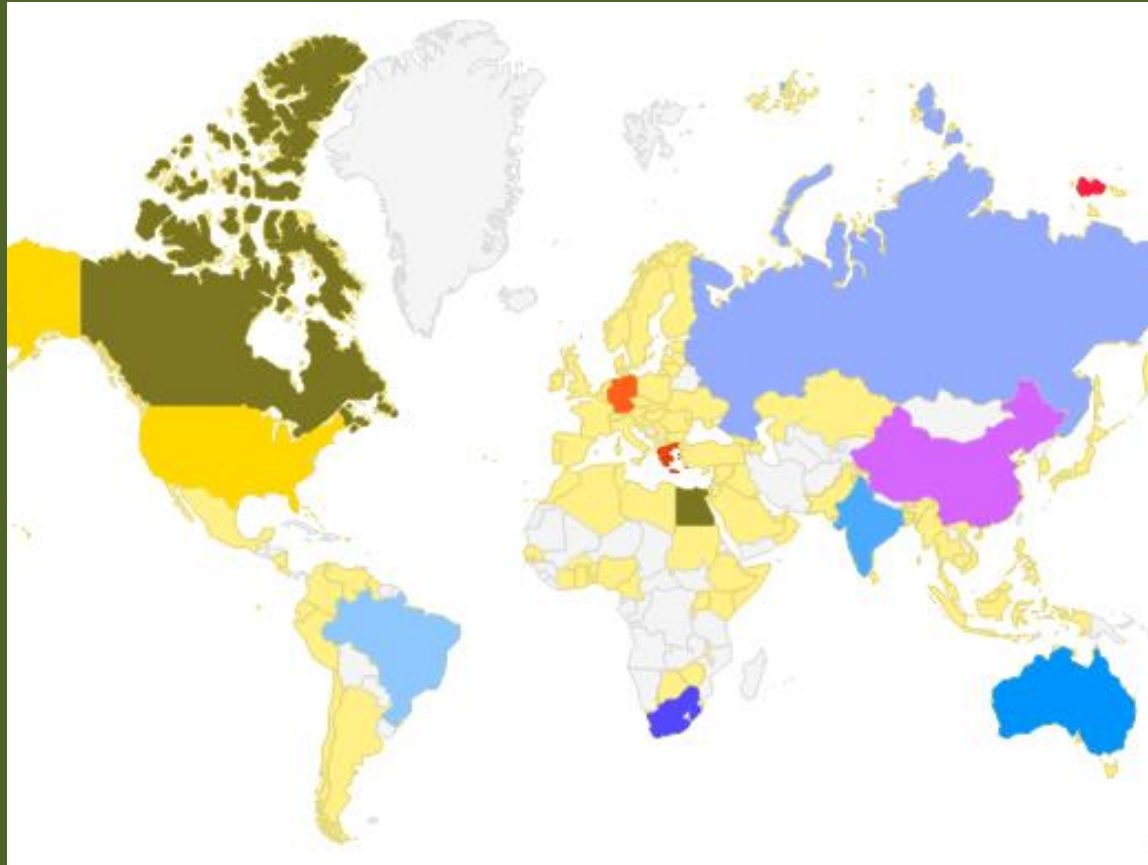


Co-funded by the
Erasmus+ Programme
of the European Union

Contents

- Descriptive Analytics
 - Correlation
 - Regression
 - Clustering
 - (A) Classification sequence

- Structured reports
 - The magic nodes
 - The BIRT environment
 - Baby steps in [structured] report design



Descriptive analytics



Co-funded by the
Erasmus+ Programme
of the European Union

Good practices

- Workspace Essentials – keep:
 - Workflow browser
 - Node browser
 - Node help
 - Console
- Workflow management
 - *Keep track of your folders*
 - Tidy-up your workflows regularly
 - *Can set up additional workspaces for ‘large’ projects*
- (Per) Workflow
 - **Name** your nodes. *Yes, now.*
 - Can copy-paste. Careful with settings though.
 - Node management: Keep only what you *really* need.
 - Save regularly.



Preparation

- Filter *out*
 - EU aggregates.
 - Non EU countries.
- Split into
 - Eurozone
 - Non-eurozone





Descriptive analytics - Correlation

- Correlation
 - Per *country*:
 - In country 'X': is there a correlation between LE, EL & GE?
 - Variables: Life expectancy vs Education vs Greenhouse emissions
 - Measurements: *Annual*
 - Per *year*:
 - In country 'X': is there a correlation between LE, EL & GE?
 - Variables: Life expectancy vs Education vs Greenhouse emissions
 - Measurements: *Countries*



Descriptive analytics - Regression

- Regression
 - Per *country*:
 - Which of LE, EL & GE interprets better the other two?
 - [Hint: 3-simple, 3-multiple regressions]
 - Per *year*:
 - For GE only. Is performance in one year indicative of the next year?
 - Emission Taxes vs Emissions:
 - Retrieve Eurostat tables:
 - taxes on emissions [env_ac_taxind2]
 - CO Emissions [env_ac_ainah_r2]
 - How does Denmark do?
 - How does Greece do?



Descriptive analytics - Clustering

- Clustering – €-zone groups
 - Hierarchical clustering on
 - Which countries are closer on LE, EL, GE?
 - Which are not?
 - How does the **distance** [!] measure affect the clusters?
 - Similarity measure: e.g. Manhattan, Euclidean, ...
 - Clustering or Linkage criteria: e.g. Min, Max, Mean, Least variance (Ward), ...
 - ‘Optimal’ no. of clusters: Elbow [variance explained], Shadow [self-similarity]
 - K-means – €-zone groups
 - How can we divide into [X] groups, based on LE, EL, GE?



Descriptive Analytics :: Classification

- Let's expand ...
 - E.g. drill down to NUTS 2 or NUTS 3 level.
 - E.g. EU **vs** the rest of the world (NUTS 1).
- *Assume the EU clusters are typical.*
 - Which NUTS2 or NUTS3 fit the clusters?
 - How do non-EU countries fit the Europe clusters?
- **Tip.** Let's get the data for the rest of the world.
 - What can we find?
- **Nodes**
 - K-means as training data
 - K-nearest neighbour as classifier



Structured Reports



Co-funded by the
Erasmus+ Programme
of the European Union

KNIME Report designer :: The BIRT environment

Preparing for the report

- The '*magic*' nodes
 - Data to report
 - Image to Report
- One [*data* | *image*] to report
 - **per table** to be used included in the report.

Workarounds

- Non-reportable items
 - Export to file from workflow
 - Import from file to report
- The Heatmap / correlation





The BIRT Environment

- **Transferrable skills**
- Data sets
- *Palette*: Report items
 - *Static*
 - Label
 - Text
 - Image
 - *Dynamic*
 - Dynamic text
 - List
 - Table
 - Chart
 - Cross tab
 - ‘*Dummy*’ aka “*Placeholders*”
 - Grid
- The logic
 - Blank Canvas
 - “Place” report item(s)
 - Test
 - Customize appearance
 - Test
 - Repeat until report design is completed.
- Various formats
 - Documents
 - PPT
 - PDF
 - HTML
 - ...



Baby steps

Get the basics right

- Box design
 - Divide page into boxes
 - Decide what goes where
 - Static & Dynamic elements
- Try / Do:
 - One table
 - One Chart
 - Descriptive labels

And move on to mastery

- Box design
 - Progressively complex layouts
 - Fine tuning properties
 - Brush up on HTML, CSS.
- Multiple report items
- Dynamic & Static elements





Baby steps

Let's do the 'excel' way.

- A report to display
 - One table
 - One chart from Workflow
 - One chart from Report designer
- Practice display settings & properties
- Let's try
 - LE data

or

 - EL data

or

 - GE data
- **ONLY.**
- *When you get it 'right', then expand to your workflow.*

Need to have

- Labels
- Data table(s)
 - Any combination of total, Females, Males
- Chart[s]
 - From workflow [image to export]
 - Dynamically created.





Now, free your inner analytics report artist, ... let go!



Summary

- Descriptive Analytics
 - Correlation
 - Regression
 - Clustering
 - Classification sequence

- Structured reports
 - The magic nodes
 - The BIRT environment
 - Baby steps in [structured] report design

