

## 1. Problem Understanding

### 1.1 Background

The Organization for Economic Cooperation and Development (OECD) is a large intergovernmental organization that is tasked with enhancing world progress, supporting international development and stimulating trade. The original impetus for the organization was to help coordinate European reconstruction efforts after World War II, by administering funding from the U.S. Marshall Plan. Since then the organization was re-founded and participation was further extended to non-European partners.

The organization is now funded by contributions from its 36 member states, and it has a yearly budget in excess of 350 million dollars a year. About a third of this budget goes to international development. Other activities involve setting tax policies, publishing, and developing open indicators of world economic activity. Most member states in the OECD are high income and highly developed countries. The total GDP of participating countries varies from 40% to 60% of world economic product, depending upon the measures that are used.

### 1.2 Decision-Maker Objectives

The OECD is organized through a set of directorates, each addressing a particular substantive area. There are directorates addressing entrepreneurship, tax policy, education and employment, innovation, the environment, governance and public affairs, statistics, trade, and agriculture. Together these directorates employ some 2500 staff members.

Your work is commissioned by the Counsellor for Strategic Foresight. The Counsellor reports directly to the Secretary General but lacks staff capacity of their own. As a result, they must raise internal awareness of significant emerging issues across the various directorates, stimulating further policy analysis and implementation.

Given the over-arching objectives of the OECD, the organization is extremely interested in addressing a range of grand societal challenges including the seven societal challenges of the European Commission, and the sustainable development goals as formulated by the United Nations. Given the specific objectives of the Counsellor for Strategic Foresight, it is important to surface new and emerging issues, and to help communicate these issues. Communication happens to the Secretary General, to other Directors, and most especially to representatives from the Member States.

### 1.3 Success Criteria

Your data analysis and accompanying visualization will be deemed successful to the Director to the extent that it:

- Involves grand societal challenges;
- Involves strategic foresight, tackling issues early in their emergence cycle;
- Is evidence-based;
- Shows causal understanding and the ability to tease out possible intervention strategies;
- Addresses fundamental uncertainties;
- Is organizationally aware of OECD missions and objectives, and of the sensitivities of the respective member states;
- Recruits additional attention, resources, and data collection from the various directorates.

Given tight human resources, and a range of significant issues to be addressed, this particular data science project will be most successful if it is a rapid survey of the particular topic.

### 1.4 Inventory of Resources

Your primary inventory for this project is to be the World Bank open data repository. This is a relatively new entrant to the world community of development analysts. The Counsellor for Strategic Foresight is interested in how this data might be better used in support of the OECD mission. The OECD is also a prominent provider of open data as well (OECD, 2018). You also have a team of up to three analytical staff members also assigned to explore your topic of choice.

### 1.5 Requirements, Assumptions & Constraints

The overall report is expected by 9 November 2018. These reports are typically 6000 words in length, and make good use of graphics, charts or maps to better communicate. The report must select one or more graphics which communicate to either decision-makers, stakeholders, or another OECD analyst. The assumption is that the report will be a first step in a further cycle of analysis as well as of policymaking.

Another requirement is that the model make good use of analytical models. The rest of the analysis staff at the OECD want to be able to utilize open data. Interest and awareness is high among analysis staff about the use of Bayesian modeling and graphical models. But they seek additional help in employing such models in their own directorates and are concerned about how best to explain causal features of the model, and how to best communicate credible ranges of uncertainty given the data.

It is accepted that the modeling project is constrained to examine only the most recent years of data. Panel data, showing changes over multiple countries over multiple years, is currently being produced elsewhere in the OECD. Working with this data is slated for further development. A further concern is that such modeling is typically a-causal – observations are made concerning relative growth or decline, without much attention to the underlying factors which drive such trends. It is your task to uncover such trends.

### 1.6 Risks & Contingency

There are several risks about which you should be aware. The primary concern lies in reverberation. The findings you make if successful will impact others outside the OECD. These outsiders will then lobby the OECD in support of their prospective concerns and interests. This process of analytical results escaping the organization, and then being reflected back is known more generally as “reverberation” (Putnam, 1988).

Reverberation takes two primary forms. Member states are sensitive to how they are portrayed in OECD publications. While they value the independent and objective advice received, they naturally wish to be portrayed in the best possible light. A second form of reverberation comes from recipients’ advocates. They naturally wish their issues and regions to receive the greatest amount of attention and aid from the OECD.

The OECD also represents the current liberal world order. It could therefore be criticized by those who think that developed countries, and their respective corporations, have done too little to mitigate the adverse impacts of economic, technological and environmental change for the underprivileged. Others criticize the OECD for its internationalism – its willingness to pursue a cross-national agenda on a range of complex issues including migration. Your best contingency plans against such criticism is to be aware of the concerns of such outsiders. Address their concerns before they are raised. You may do this in either of two ways: by acknowledging the legitimacy of these arguments, or by posing appropriate counterarguments within your report.

### 1.7 Terminology

You may encounter the following institution or domain specific concepts while writing your report.

- *Development Aid Committee*. Those member states contributing directly to the international aid budget.
- *Directorate*. One of the major bureaus of the OECD, centered around a long-term and functional goal of the OECD.
- *Grand societal challenges*. Grand societal challenges are a form of mission-oriented funding which build enthusiasm from both funders and researchers to work on common problems.
- *Graphical models*. A style of machine learning which grew out of classical statistics and encompasses elements of Bayesian statistics and deep learning.
- *Member State*. One of the thirty-six participating nations in the OECD.
- *Secretary General*. The political appointee who heads the OECD.

## 1.8 Costs & Benefits

The costs of the report mostly concern personnel costs. Your team of analysts are expected to spend 150 hours together on preparing this report. Given salary costs, social costs, Parisian cost-of-living adjust and overhead the total cost of the report is 15k Euros. An additional 3500 Euros have been allocated for dissemination, including social media engagement, web pages, press reports, and report printing.

The benefits of the report are mostly non-monetary. However, if the report improved development funding by even one-half of one percent, the monetary value of the report could exceed 1.75 million euros. Regardless the additional value of the report stems from its ability to support the OECD in evidence-based policymaking, and the capacity of the report in helping the OECD take the lead in important policy discussions.

## 1.9 Data Science Goals

The data science goals of the project stem directly from what is written above. These criteria are more technical in character. The data science project should

- Use open data sources
- Utilize open software resources (in this case the R language and R Studio)
- Demonstrate the possible relevance of a graphical modeling approach
- Identify further data for testing
- Identify further credible hypotheses for investigation

## 1.10 Data Science Success Criteria

The criteria for project success are given in the form of the following table or rubric.

	5 (or less)	6	7	8	9 (or more)
Task 1. Problem Understanding	A problem is not selected; the data and graphics are not adapted to the needs of the decision-maker	A problem is selected, but poorly described	A problem is compellingly described. Links are made by the appropriate selection of data or the appropriate selection of graphics	A problem is compellingly described. Both data and graphics are appropriately used.	A problem is compellingly described Both data and graphics are compellingly used.
Task 2. Graphics and Visualisation	Not presented	Charts are presented using base R	Charts are presented and customized; the insight from the chart is described in text	Custom graphics are included using ggplot or other packages; the graphics are tailored to for decision-makers, analysts or stakeholders	Multiple effective graphics are shown throughout the report

	5 (or less)	6	7	8	9 (or more)
Task 3. Data and Preparation	Data was not selected or prepared.	Data is selected, but its appropriateness for the problem is not discussed.	Data is selected and appropriately used. Inappropriate assumptions about the data are introduced.	Appropriate cleaning was performed and discussed. This is fully evidenced using the CRISP-DM reporting mechanisms.	Exceptional efforts were made to collect additional data or to fuse multiple data sets.
Task 4 Analysis	A graphical model in R is not presented.	The model selected is not discussed	The model selected is discussed	The model selected is strongly motivated; the Bayesian assumptions are fully discussed	The model selected is strongly motivated and additional insight is provided
Task 5. Reporting	The report is incomplete.	Discrepancies or limitations of reporting given the CRISP-DM framework are discussed.	The reporting is complete; the CRISP-DM framework is used in a productive manner.	The reporting is both concise and complete.	The report invites clear next steps for additional analysis or decision-making.

### 1.11 Project Plan

The following plan is suggested for your teamwork.

Date	Phase	Task	Description
Tue, 15 October 2019	Problem Understanding	Case	Case Selection, Select an appropriate case
Wed, 16 October 2019		Graphic (optional)	Create a graphic clarifying the situation.
Fri, 18 October 2019	Data Understanding	Design (optional)	Select an appropriate database
Mon, 21 October 2019		Data	Select appropriate variables
Wed, 23 October 2019		Reporting	Complete the data understanding report
Thurs, 24 October 2019		Graphic	Create graphics showing key relationships in the data
Monday, 28 October 2019	Data Preparation	Data	Prepare the data
Tuesday, 29 October 2019		Data (optional)	Enrich the data in light of exploration
Wednesday, 30 October 2019		Reporting	Complete the data preparation section
Thursday, 31 October 2019	Analysis	Modeling	Select and run an appropriate graphical model
Friday, 1 November 2019		Modeling (optional)	Elaborate on your initial model
Monday, 4 November 2019		Reporting	Complete the analysis section
Tue, 5 November 2019		Graphics (optional)	Explore the posterior predictive distribution
Tue, 5 November 2019			Final report due

This is included only in an advisory manner, and because it is a standard part of the CRISP-DM framework. No interim deliverables are required but the final. In the event that you are working in a team the final report should have a statement, indicating there was shared or delegated responsibility for tasks including modeling, reporting, graphics, case design and reporting.

#### References

OECD (2018), "OECD Data: Find and share the latest OECD data," Accessed online 18 October: <https://data.oecd.org/>.

Putnam, R. D. (1988). "Diplomacy and Domestic Politics: The Logic of Two-Level Games," *International Organization*, vol. 42, no. 3, pp. 427-460.