

Module name: Data Analytics Concepts

Assessment: Coursework

Course dates: 5-13 November 2018

Objectives

- Produce a briefing summarising key information from HM Treasury's National Infrastructure Pipeline.
- Use Excel to clean up, analyse and create visualisations out of the data for use in your briefing.
- Provide a clear audit trail of every calculation and clean-up you have performed.

Assignment description

You are provided with a dataset published by the UK government which outlines its 2017 infrastructure 'pipeline' – all national infrastructure programmes currently under construction or scheduled to commence construction.

This dataset has been modified manually in some areas to randomise numerical values and introduce additional errors.

Each row in the dataset represents a different project/programme.

You will take this dataset and use the skills you have developed on this course to explore it, perform clean-up where there are inconsistencies, analyse its findings and report on the key themes you have discovered.

You will write a short briefing to summarise in words and charts what you have found, submit an Excel workbook with your calculations, and an audit file listing what you have done.

Assignment tasks

Submit three documents for assessment:

- a word-processed briefing, summarising the key findings in short sentences/bullet points and charts
- an Excel spreadsheet containing all of your working and charts
- an audit file listing the modifications you have made to the dataset. This can be in .txt. form or word-processed

Task one

Take the spreadsheet you are provided with, read and understand it, and make changes to it. As part of this you should:

- 1) Start a .txt or word-processed file and maintain a list of the key changes you make to the dataset. These do not need to list all the formulas actually used, but describe where you have made a change, and which function you have used to make it. Examples might be:

1. Created derived column (F) using VALUE() to resolve non-numerical entries
 2. Created derived column (G) using VLOOKUP() on population data, using region names as lookup values
 3. Created pivot table (sheet1) to tabulate total capital spending by sector and sub-sector
 4. Created line chart of data (sheet 2) in table P2:R10 from tab Autumn_2017_Pipeline
-
- 2) Correct or resolve any errors or inconsistencies you find that are **relevant** to the analysis you want to perform (specified below). You should create new columns for the corrected entries rather than overwriting the existing columns so that your corrections are clear to see. You won't be given extra credit for resolving additional errors that don't affect your actual analysis.
 - 3) Using data analysis, generate answers to the following questions within the spreadsheet. Where appropriate, the answers should take the form of a table, pivot table, or a set of new columns in the dataset.
 - a. How much capital spending, in total, is being spent on each Sector of the economy?
 - b. How does the above break down by Funding Source?
 - c. Show the average proportion of spending on projects in each UK region that comes from public sources (Capex cost publicly funded as a proportion of total Capex cost). Exclude broad geographies above regional level, such as UK, England and Wales, International.
 - d. Show how UK regions (the same geographies as above) compare in terms of **Publicly Funded capital spending on Transport** per head of population.
 - e. Convert any nominal spending figures that are in the dataset into real-terms, **using 2018/19 as the base year**. For any figures already in real-terms, convert these into 2018/19 prices.
 - 4) Create charts to visualise your findings, saving them in the worksheet.

You can use pivot tables or pivot charts to help with your analysis, but these are not required if you prefer to create tables and charts manually.

You **must not** delete or modify any of the original data in the spreadsheet. All of your calculations and clean-up work should use new columns you have created based on the original data. The original data should be there for the examiner to see, alongside your own work.

Part of the marking is based on how well you demonstrate your knowledge of Excel formulas. Try whenever possible to use a formula instead of making manual entries – for example, when correcting errors (this won't always be possible).

Task two

Produce a short briefing in Word format of **no more than three sides of A4**, summarising your key findings. You should draft it as if the audience is a group of civil servants and economists from the Treasury interested in learning about the UK's infrastructure spending.

As part of this you should include:

- 1) In brief bullet points at the start, summarise the most interesting facts you have learned about infrastructure spending from the questions you answered (questions a-e in task one). You should judge what are the most interesting, important and

relevant points to go here. You **don't have to** have a bullet point about **all** of the questions a-e.

- 2) Insert some of the key visualisations you have created using the data, and add short paragraphs to each explaining what they are showing. Again, you don't have to have a chart responding to all of the questions set.

The briefing is your opportunity to expand upon the questions set in task one. You will be marked on **how well you draw out key, relevant stories from the data**. Don't just list the answers to each of the questions set in task one.

You don't need to write very much text. The key is to summarise your findings in an engaging way, using bullet points, short paragraphs and visualisations.

Do not go over the three-page limit unless you have a good reason for doing so (such as a large visualisation).

Marking scheme and structure (total of 100 marks)

Spreadsheet

Data clean-up (15 marks)

15 marks – All relevant errors and inconsistencies have been cleaned up using efficient formulas

10 marks – Some errors have not been cleaned up, or the formulas used are long-winded or incorrect

5 marks – Important errors from the data have not been cleaned up and materially affect the analysis

0 marks – More than one important error has not been cleaned up.

Poor clean-up can in turn affect accuracy marks in other criteria

Analysis – questions answered (50 marks, 10 per question)

10 marks – Question answered correctly using relevant and efficient formulas

5 marks – Minor inaccuracy or long-windedness in formulas

0 marks – Significant inaccuracy

Visualisations – accuracy (15 marks)

15 marks – All charts are accurate

10 marks – There are some inaccuracies or minor errors in the charts

5 marks – There is a significant inaccuracy (which materially affects a reader's understanding of the data) or small inaccuracies are common

0 marks – There is more than one significant inaccuracy

Audit trailing (10 marks)

10 marks – All your calculations are clear and easy to follow from the audit file

5 marks – Some calculations relevant to the analysis aren't obvious or haven't been shown

0 marks – Several calculations are missing or are difficult to follow

Formulas should be visible to the examiner from the spreadsheet as well – the audit file is not a substitute

Briefing

Headline bullet points – relevance and engagement (5 marks)

Marked on ability to communicate relevant points with brevity

Visualisations – engagement (5 marks)

Marked on how easy the charts are to follow and marked down for elements that are unclear or potentially misleading