Government Statistical Service

Data Science competencies

Statistics

EO Level - Theoretical understanding of basic statistics – *for example*, correlation, standard distributions, or regression. But, only minimal applied **execution** examples – *for example* university coursework with an analytical technique prespecified.

HEO Level – EO level + Can explain aspects of inferential statistics to a lay person, and communicate uncertainties around analytical evidence. Has **autonomously designed** analysis to support a decision. Examples *might* include: core statistical techniques for model selection, hypothesis testing, or time series forecasting.

SEO Level – HEO level + Can justify and communicate various **innovative** analysis examples, as stories intuitive to their audience. Additional examples *might* include supervised / unsupervised machine learning for complex data sets, or Natural Language Processing to structure data for analysis.

Data

EO Level - Theoretical understanding of data preparation – e.g. standardising formats across sources, dealing with missing values, using semi-structured data formats (e.g. XML/JSON), or geospatial (e.g. GIS). But, only minimal applied **execution** examples – e.g. university coursework with a dataset provided.

HEO Level – EO Level + Has **autonomously designed** data processing to support analysis. Examples *might* include: imputation; using metadata; database design; applying sampling methodologies (e.g. random, stratified, cluster, quota). Understands Open Data's limitations in terms of representativeness. Has clearly considered data ethics (e.g. robots.txt).

SEO Level – HEO Level + Can justify an **innovative** data processing example. Examples *might* include imputation to preserve representativeness; efficient storage/access through noSQL database structure (e.g. graph, document, geospatial); structuring data from APIs, web scraping; or, judicious assessment of data – its suitability, canonicity etc.

Programme/Computing

EO Level - Theoretical understanding of computing – for example, can name different languages and tools (e.g. Python, Java, SAS) and express the difference between storage and memory. But, only minimal applied **execution** examples – e.g. university programming coursework with prespecified objectives

HEO Level – EO Level + Can explain programming to a lay person, and justify using Big Data technology from a hardware perspective. Has **autonomously designed** programs and systems. Examples *might* include designing a program for a new problem AND writing OS scripts (e.g. BASH), using data transfer (e.g. HTTP/SSH), or using version control (e.g. Git). Has clear experience of unit testing.

SEO Level – HEO Level + Can justify an **innovative** applied computing example. Examples *might* include: efficient programming practices (e.g. vectorised operations, fast search / sort / matching methods); numerical stability of algorithms; containerisation (e.g. Docker) and setting up a process pipeline for reproducibility; or, developing a web application to accurately present information.