DEFINITION OF ANALYTICS PROJECT ATTRIBUTES

The Table below shows accumulated analytic project attributes that relate to analytics projects, the potential values that these attributes can take, the literature sources that discuss the application of said potential values (it should be noted that the columns with "-" means that these potential attributes were defined by the author), and the last column contains comments regarding these attributes.

List of Analytics Project Attributes

Project Phase	Attributes	Potential(example) Values of Attributes	Literature source	Comments on the Attribute
Initation	Project Context/Field	Industrial Process	(Runkler, 2020)	This attribute of the analytics project refers to the problem domain within which the data analytics project is being carried out.
		Business		
		Biomedical		
	Object/System of intrest	Organisational unit	-	This attribute defines what object(s) or system(s) the analytics project is concerned with.
		Energy Systems	-	
		Computer systems	-	
	Analytics type	Descriptive analytics	(Cote, 2021)	Defining the type of analytics that is being carried out within the project is very important, especially for the analysis phase of the project since the goal of the project is very dependant on the type of analytics.
		Diagnostic analytics		
		Prescriptive analytics		
		Predictive analytics		
	Project Documentation	Project Plan	-	The documentation that is provided can change the amount of detail an analyst is provided on things that relate to company data, available resources and expected results.
		Data Flow Diagram	(Olayan, Patu et al., 2013)	
Acquisition	Data type used	Numeric data	-	

Project Phase	Attributes	Potential(example) Values of Attributes	Literature source	Comments on the Attribute
		Textual data	-	The types of data that are used with any analytics project can be placed into one of these categories, regardless of the format. The type of data used within a project can affect factors such as the type of analytics method that will be used as well as the possible visualisation methods.
		Graphical data	-	
	Dataset(s) used	Preexisting dataset	-	Whether or not the datasets used in the analytics project are already present or need creating changes the work required within the data acquisition phase. The types of data that are used with any analytics project can be placed into one of these categories, regardless of the format. The type of data used within a project can affect factors such as the type of analytics method that will be used as well as the possible visualisation methods.
		Created dataset	-	
	Data source(s)	System(s)	-	The source from which data is collected affects the data extraction methods that are required as well as how external factors such as data privacy laws affect the analytics project.
		Data stores(s)	-	
		Personal(humans)	-	
	Dataset cardinality	Single-source	-	This attribute within an analytics project defines whether or not the analytics project uses a singular data source or multiple sources. Depending on this variable the complexity of the extraction process changes. This is derived from the mathematical term ‘cardinality of sets’.
		Multi-source	-	
	Data extraction technique	Web data extraction techniques i.e Roadrunner	(Salah, Okush et el., 2019)	The technique used to extract data is dependent on the form of data that is extracted and who or what said data is extracted from. This makes the data extraction process a large part of the data analytics project.
		ETL(Extract,Load, Transfer)	(Kadadi, Agrawal et al.,2014)	
		Questionnaire		

Project Phase	Attributes	Potential(example) Values of Attributes	Literature source	Comments on the Attribute
	Data Integration	NOSQL Algorithm	(Kadadi, Agrawal et al.,2014)	Data integration is dependent on the cardinality of the dataset. If there are multiple sources of data then data integration might be necessary inorder to combine them into a singular format.
Analysis	Analysis technique/ method/methodology	Correlation Analysis	(Runkler, 2020)	The technique of analysis being done impacts most aspects of the analytics project and vice versa. Therefore the selection of the techniques used in the project and the definition of technique used is essential for the success of the project.
		Regression Analysis		
		Text Analysis	(Gururajan, Clark et al., 2014)	
		Clustering	(Runkler, 2020)	
	Analysis tools/software	Pandas(library) in Python 3(Programming Language)	(Unpingco,2021)	The tools that are used to carry out the analysis change the nature of the analytics project similar to the selected analysis technique. A good demonstration of this is how the software will affect "Data types used" because different tools have different inbuilt data types, as well as rules regarding how those data types can be manipulated.
		R(programming language)	(Ghahramani & Prokofieva, 2021)	
		Hadoop	(Praveena & Bharathi, 2017)	
Presenting of results	Report Specification	Results formating	-	Most if not all analytics projects produce a report that is given to the clients as a project artefact. Therefore formatting this report in the most acceptable way is vital to client satisfaction.
		Catergorisation of results	-	
	Graphical results	Line graph	-	The visualizations used within a data analytics project are not the most vital characteristic but they determine how well the finding are presented, therefore it is the culmination of the project. It is dependent on the characters within the acquisition and analysis phase.
		Bar Chart	-	
		Histogram	-	
	Interative results	Tableau dashboard	(Praveena & Bharathi, 2017)	Through the use of various software interactive tools such as BI dashboards can be produced as a

Project Phase	Attributes	Potential(example) Values of Attributes	Literature source	Comments on the Attribute
				result of the analytics, one of the most popular visualisation tools is Tableau.

References

Runkler, T. A. (2020). Data Analytics. *Springer EBooks*. https://doi.org/10.1007/978-3-658-29779-4

Cote, C. (2021, October 19). 4 Types of Data Analytics to Improve Decision-Making. Business Insights Blog. https://online.hbs.edu/blog/post/types-of-data-analysis.

Olayan, N.,Patu, V.,Matsuno, Y., Yamamoto, S., A Dependability Assurance Method Based on Data Flow Diagram (DFD), 2013 European Modelling Symposium, Manchester, UK, 2013, pp. 113-118, doi: 10.1109/EMS.2013.20.

Salah, M., Okush, B. A., Rifaee, M. A., A Comparison of Web Data Extraction Techniques, 2019 IEEE Jordan International Joint Conference on Electrical Engineering and Information Technology (JEEIT), Amman, Jordan, 2019, pp. 785-789, doi: 10.1109/JEEIT.2019.8717519.

Kadadi, A., Agrawal, V., Nyamful, C., Atiq, R.,Challenges of data integration and interoperability in big data, 2014 IEEE International Conference on Big Data (Big Data), Washington, DC, USA, 2014, pp. 38-40, doi: 10.1109/BigData.2014.7004486.

Gururajan, R., Clark, K., Moller, S. et al., Reliability of Qualitative Data Using Text Analysis - A Queensland Health Case Study, 2014 3rd International Conference on Eco-friendly Computing and Communication Systems, Mangalore, India, 2014, pp. 303-308, doi: 10.1109/Eco-friendly.2014.68.

Unpingco, J. (2021). *Python Programming for Data Analysis*. Springer EBooks. https://doi.org/10.1007/978-3-030-68952-0

Ghahramani, A., Prokofieva, M.,Visualisation for social media analytics: landscape of R packages, 2021 25th International Conference Information Visualisation (IV), Sydney, Australia, 2021, pp. 218-222, doi: 10.1109/IV53921.2021.00042.

Praveena, M. D. A. , Bharathi, B., A survey paper on big data analytics, 2017 International Conference on Information Communication and Embedded Systems (ICICES), Chennai, India, 2017, pp. 1-9, doi: 10.1109/ICICES.2017.8070723.