CONTRIBUTION DOCUMENT DAIA PYTHONATIC

Layout-Eva

Introduction-Antonia

Data requirements

* Stakeholders – Antonia/Eva
* Identify required data elements – Bas, Eva rewrote it?
* Identify Candidate data sources - Eva
* Data quality criteria – Antonia
* Overview of the data requirements – Antonia and Eva rewrote it

Data collection

* Determine what information you want to collect - Eva
* Where to store your retrieved data – Ryan, bas rewrote it and Eva added stuff
* Use a (traceable) version and naming system – Ryan, Eva rewrote it
* Determine how and how often you want to retrieve (or reload) the data- Bas
* Data extraction – David/Alex/Eva
* Conclusion - Antonia

Data Understanding/Data preparation

* Structuring the document – Antonia (With a bit of help of Eva) (didn’t save)
* Data sourcing – Eva (but was deleted in the final version because it’s partly in the Data Quality conclusion and the other part in the document)
* Data Quality Analysis Conclusion – David
* Data understanding (supposed to be data cleaning)- Eva (with a bit of help of David)
* Analytic approach
* Data understanding and visualisation
  + Delay and max numbers of stops – Eva
  + Total train delay – Antonia
  + Average train delay per day – Antonia
  + Percent train delays - Alex
  + Amount of train delays - BAS
  + Correlation investigation - David
  + Dealing with missing values - Alex
  + Time groups - Alex
  + Balance between delays and no delays – Alex
  + Negative train delays - Alex
  + Conclusion – Antonia
  + Solve errors that exist because the of merging of different notebooks – Antonia/Alex (a bit of help from Eva) (Didn’t save)
  + Risks – Alex with a bit of help of Eva (we send this in the mail because it didn’t save)