GROUP ASSIGNMENT

Description

The city of New York shares many datasets to the public under the open data initiative. You are CTO of openDataNews. The goal of open data news is to build services and products on top of data that is available through open data initiatives that can generate revenue for the company. You now have been asked to create a working prototype of a product around the taxi trip dataset of NY in Tableau.

Data

A starting data set can be found at https://data.cityofnewyork.us/Transportation/2016-Green-Taxi-Trip-Data/hvrh-b6nb/data. Use other sources from Open Data New York to enrich or support your business case.

Deliverables

A tableau that works correctly on a computer that does not have the dataset available. So when I open it, it has to function correctly. This means if you create a database, it has to be accessible for the teacher as well.

A presentation will be given on the day of the exams. The powerpoint needs to be handed in on the 20th of December as well.

Deadline and submission guidelines:

By the 20th December 2017 at 23:59 CET, every group member has to upload every document to Moodle.

A penalty of 5 points out of 20 is used for every day you are late. From this follows, you can not pass the group assignment when you are two days late.

Grading Rubric

Criteria	Weight	0	1	2	3	4
Data Sources	15%	Only the data source provided is used		A few data sets are used. All of them are relevant.		Many data sets are used, all of them add value to the business case.
Data Modelling	30%	The data is prepared incorrectly. Most columns contain wrong data types. Incorrect use of dimensions and measures	There is no relational model but the majority of columns have a correct data type and have been assigned correctly as a measure or a dimension.	The data is not put in a relational model but all the columns have been assigned the right data type. Every column is defined correctly as a measure or a dimension.	The data is put in a relational model outside tableau, most of the columns are modelled correctly. However the data model contains some errors.	A correct data model such as the relational model is used. Every column has been assigned a proper data type. Data is stored in an efficient way that enables fast querying.
Visualizations	30%	Few visualizations are used and they are very simple. The visualizations contain errors such as overinking, distortions of the axis, …	Only a few and simple data visualizations are used but the visualizations are done correctly.	There are sufficient but simple data visualizations that have been executed correctly.	There are many data visualizations from which some require special techniques such as double axis. At least one dynamic visualization is used.	At least as good as 3 but there are also data visualizations that require a special data preparation specifically for this graph. (for instance a network graph)
Business Analysis	15%	Visualizations are not in function of the business case		Visualizations support the business case		Visualizations support the business case. Management tools such as what-if analysis, sensitivity analysis, … are used to gain deeper insight.
Presentation	10%	The audience doesn't get any closer to solving the business case after hearing the presentation.	A data driven presentation but there are some inconsistencies in the story.	A data driven presentation. After the presentation, the audience gets actionable insights	A compelling story that is well supported by data driven evidence. After the presentation, the	A compelling story with excellent delivery that is well supported by data driven evidence. After the

			audience gets actionable insights	presentation, the audience gets actionable insights.