Domain: say we have an ecommerce site with products divided into categories like toys, electronics etc. We receive events like product was seen (impression), product page was opened, product was purchased etc.

Task #1:

Enrich incoming data with user sessions. Definition of a session: for each user, it contains consecutive events that belong to a single category and are not more than 5 minutes away from each other. Output should look like this (session columns are in bold):

eventTime, eventType, category, userId, …, **sessionId, sessionStartTime, sessionEndTime**

Implement it using 1) sql window functions and 2) Spark aggregator.

Task #2:

Compute the following statistics:

* For each category find median session duration
* For each category find # of unique users spending less than 1 min, 1 to 5 mins and more than 5 mins
* For each category find top 10 products ranked by time spent by users on product pages - this may require different type of sessions. For this particular task, session lasts until the user is looking at particular product. When particular user switches to another product the new session starts.

General notes:

* Ideally tasks should be implemented using pure SQL on top of Spark DataFrame API.
* Spark version 2.2 or higher
* README file in the project can be a plus
* [Example of the input file](https://drive.google.com/file/d/1bq4dH1VXOcLymf7fUrWKF_zWwZsbGj8x/view?usp=sharing)