Project Name: TBD

Project Proposal:

Development of an analytics platform derived from public data available on OpenPayments.com. Due to Sunshine Act regulations, physicians must report money received from interests, ie drug and medical device companies. These payment records are publicly available and can be potentially searched and analyzed for business and market intelligence purposes. The analytics platform would primarily support business development and market research work.

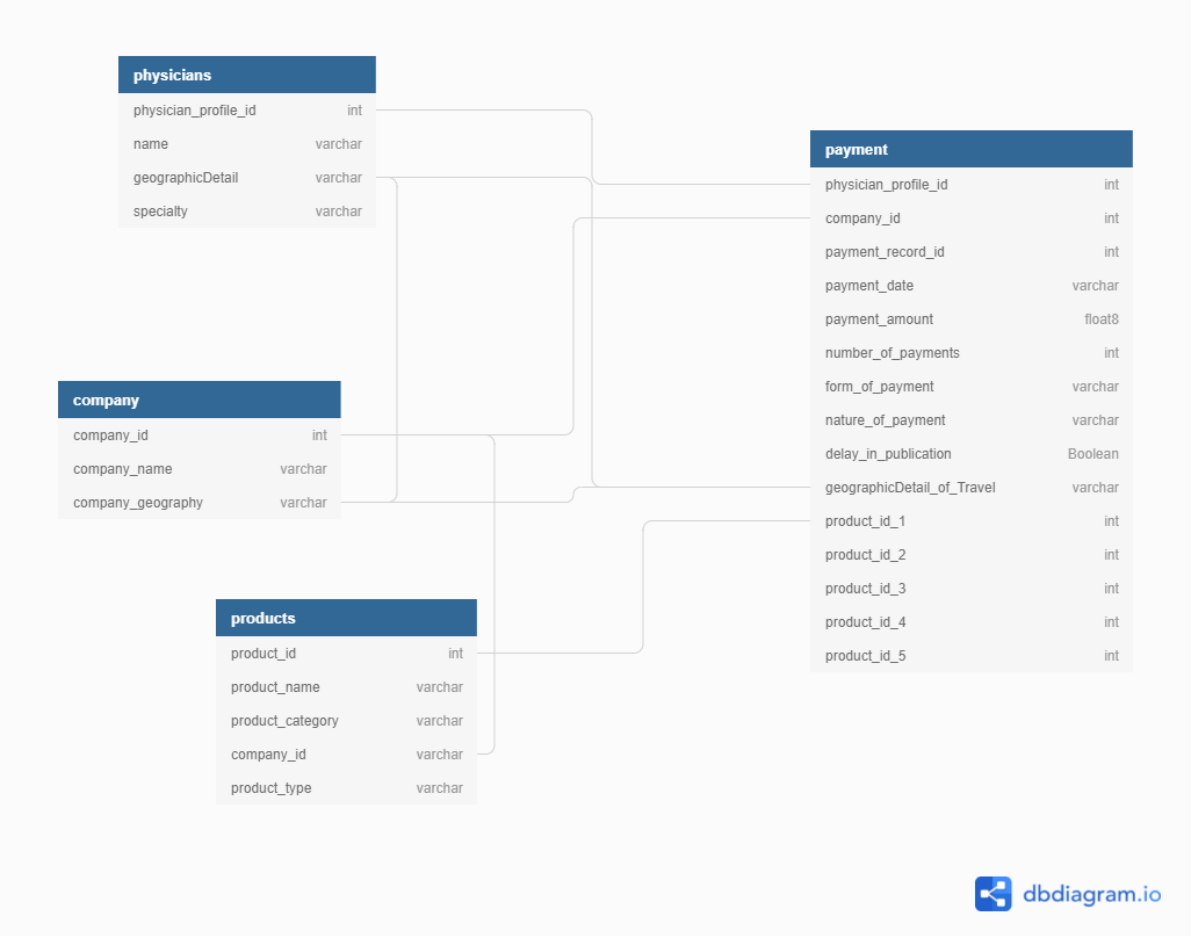
Scope of Work:

Work can be divided into two parts: Self-Contained (SC), and External(E).

Self-Contained work relies on data only from OpenPayments.com.

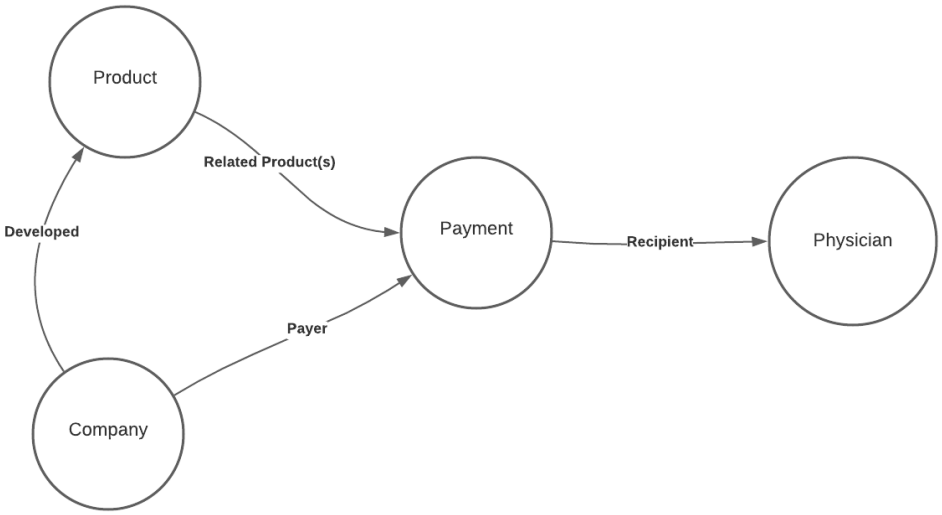
External work requires integrations with either other data sources, or use of other services.

1) Acquire Data (SC):
    a) Data Download from OpenPayments.com, full set for all years:
    b) https://www.cms.gov/OpenPayments/Data/Dataset-Downloads
2) Extract Data and Build Database (SC):
    a) Clean Data: primarily making sure product names, city/states are consistently capitalized and formatted. Physician and Companies are already identified by unique IDs.
    b) Transfer data from the excel files into database, either relational or graph.
        i) Open to discussion about potential structure, and details to capture
        ii) Proposed Structure if Relational:
        iii) Payments captured in main table, views created by pulling in data from physician, product, and company tables.



Proposed Structure if Graph:

iv) Four main node types, one for each payment record, one for physicians, one for companies, and one for products. Details captured would be similar to above tables.



3) Analytics (SC):
   a) Backend development of data pull from database to support analysis items listed below and more, export and create physician and company profiles, run PageRank/node similarity(if using graph db), etc. Known algorithms like PageRank shouldn't require work to implement, ie, Neo4J has built-in capability, but does require prep of the graph.
   b) Backend development of analytics presentation-summaries, graphs, tables, visualizations, etc
   c) Proposed analytics platform capabilities:

      i) General Market Analysis-mapping of what are the biggest players in a specific field, what products, and who they are spending their money on. Allow for a bird's eye view of the state of a market, and potential changes over time.
      ii) Company Analysis-creation of company profiles- what products they produce, which are their likely priorities (ie. large amounts of money spent on it relative to the market and other products, etc) and who their top paid physicians are.
      iii) Individual Analysis-identification of persons of interest. For example, people that receive large amounts of money from companies, those involved in lots of different products, etc. Identification of which products they are involved and for what reasons (consulting, royalty, travel, etc). Allow for granular subsetting by physician specialty, location, products involved in, etc. Can also be used to identify hospitals (see last on the list below).

   d) More specifically, Individual Analysis would allow for the following:
      i) Travel history identification-granular analysis of who is traveling where and when. Place of travel is usually listed on payments. Intelligence can be gathered based on a number of factors, ie large amounts of people traveling to a certain city, travel to smaller, out of the way cities, etc. Can also ID cities of interest for a company, ie large amounts of money spent to send people there. Geographic display of travel could be created as well.
      ii) Consulting and Royalty payments- granular analysis of who gets paid for consulting and royalties and for what products. These are likely physicians with major involvement in the development of a product, either a main inventor or a main consultant. These would then be ideal candidates for networking when developing a product and getting feedback, as they would likely have been involved and have knowledge of the process/setbacks/challenges that they went through, and how the market responded.
      iii) Education Payments-represents likely users of products, or those that have likely physically tried a product. Identification of actual user of a product out in the field, and therefore a good candidate to solicit feedback on a product during market research. Can be cross referenced with other analytical criteria to screen for "Organic Growth" candidates, those who were not intimately involved and therefore were attracted to the product for some reason. Follow up can help establish reasons for product selection over other competitors and improve market understanding.
      iv) Hospitals: Hospitals are also listed if they receive payment, equipment, etc from a company. Combined with research payments, can be used to identify centers of activity for a given company.

This does not encompass all possible analyses, but just the immediate ones identified as plausible to implement. As more analytical work is performed on the data and feedback gathered, more possibilities may arise and be implemented.

4) FrontEnd (E):
    a) User interface, web, app, or other, to query the database and bring up analytics.
5) Integration (E):
    a) Integrate physician profiles and payment data with the following data sources
        i) Integration with the Patent Databases.
        ii) Integration with the Citations Database (Cartographer)
        iii) Integration with Clinical Trails.gov data
        iv) Integration with NPI Registry.

Sections 1,2,3 are priority for development, and are core to the project.

Section 4, FrontEnd, is a nice-to-have, and would be done if potential users requested it.

Section 5, Integration work, is a nice-to-have, and would be performed after the core is in a stable state. The citations database requires more development, the other databases listed would require more data acquisition, cleaning, and cross-referencing.

Validation:

- User Testing by app or analysis request
- Creation of ‘demo’ report on an individual or company possible with the platform