

The link to the raw data for this exercise can be found [here](#).

SQL (SQL Server)

1. In the provided dataset (located in Google Sheets, but feel free to export), there are several sheets representing data from different companies and one sheet entitled "Fact Table". Imagine that each company sheet is a unique table in a SQL environment and write a query to combine the data from those into the format of the Fact Table.
2. Below is a sample SQL query of the test database (tdb) using the Microsoft SQL Server syntax. It is currently not working. Can you debug it?

```
select plat.Platform, rev.[Product] as prod, sum(rev.Revenue), sum.(rev.[Units]),  
reps.[Sales Representative] as reps
```

```
from tdb.[Fact_Table] as rev  
left join tdb.platform as plat  
on rev.[Platform ID] = plat.[ID]  
left join tdb.[Sales Rep.] as reps  
on plat.[Account Manager ID] = reps.[ID]
```

```
where prod is not null and  
[Environment] not like '%unknown%'  
and Revenue > 0
```

```
order by Revenue desc
```

```
group by plat.Platform, reps.reps
```

Python

1. In the provided dataset, there are several sheets representing data from different companies. Please provide visualizations showing revenue trends across the full set of clients, products, devices, and formats over time.
 - a. **For this question, please feel free to use whatever Python library you would like.**

2. Suppose these companies are our clients. How would you go about deciding which products to pitch to which clients? How would your answer change if there were 1000 clients in the dataset?
3. Assume that each active client provides new data to us, in the format given in their particular tab of the spreadsheet, on a monthly basis. How would you build an aggregated dataset? Which fields would you include, which would you allow to be nullable, and which would be non-nullable?

BI Data Modeling

- a. If the data in the Fact Table were viewed in a BI visualization broken out by "Media Type" from the Media Type table, why would the revenue total not match the sum of the "Revenue" column in the Fact Table?
- b. Using the notion of Parent-Child hierarchies, how would you structure the data and create a measure that checks whether a Platform is assigned to a specific sales rep. or a subordinate of theirs?
- c. In our data model, the Platform table is joined to the Sales Rep. table on "Account Manager ID" and "ID", respectively. Imagine applying the above measure to a filter on a visualization so it only shows data belonging to a rep. or their team. This visualization shows nothing when "Account Manager" is a field, but does populate correctly when "Sales Representative" is a field. What does this say about the relationship between the Platform table and the Sales Rep. table?