

CAPSTONE PROJECT DATA QUALIFIERS

OBJECTIVE The following guidelines support selecting data suitable for your Capstone Project. Considering these may prevent common pitfalls and ensure enough depth for meaningful analysis. An Analyst's priority and concern for a subject matter is often reflected in the quality of the questions asked and insights offered.

Consider a current business challenge you're working on. Pull a set of data that can inform your business challenge and form a strategic recommendation using that data. How would you visualize this data and gain buy in for your recommendation from your brand or functional president?

Examples of business challenges: increase retention, collect more names, increase revenue from our existing users, win new users from a new segment, increase e-comm sales for your brand.

Focus on a few things. Expand as time allows. Don't boiling the ocean!

PRIMARY DATA PRIMARY DATA TABLES AND INFORMATION

- Recommended quantity: > 10K but < 1M rows
- Approximately 12 - 20 columns
- 2/3 dimensions and 1/3 measures
- Dimensions include time, location, characteristics about the data (Product sales data may include order date, return date, categories, sub-categories, vendors, customers, shipping methods, sales person, territories; whereas, hospital data may include dates and times of operation, departments, specialties, ratings, etc.)
- One of the dimensions should be a unique identifier – which may be synthetic (created or combined existing fields) or natural (identification numbers – order id, customer id, serial numbers, etc.)

SECONDARY DATA REFERENCE TABLES OR SECONDARY SUPPORT INFORMATION

- Any size table but must be connected to primary data via a key or unique identifier.
- The purpose is to enrich the primary data.
- Include details or reference data like locations demographics or statistics, customer contact information, product components, etc.
- Reference examples for sales data are customer contact details, territory demographics, store location address, or components of a product.
- If a secondary data source is not available, Tableau's built in demographic data, available within Map Layers, can be connected to user data by various standard geographic data elements (i.e., zip code, county, etc.).

ANALYTICS IS YOUR DATA ANALYTICALLY FRIENDLY?

READY?

- Avoid summary or pre-aggregated data. Does the level of aggregation exceed the level of detail required in the analysis?
- Is aggregation level consistent throughout? (ex: summary data vs. granular detail?)
- Column names should be in the first row only. (ex: Double stacked headers create additional cleaning work).
- Data should be in columnar format. Row-based data will likely need to be pivoted.
- Survey data is not ideal. If used, combine with additional reference data about the topic or population. This could include demographics of the surveyed population.
- When possible, use informative human readable field names.
- Use standard delimiter types (i.e. spaces are not a good delimiter).
- Split pre-concatenated fields.