**Best Practices for Data Storytelling**

So, in the video, Utkalika discusses the following golden rules while delivering a presentation for a story backed with data.

1. **Present insights, not data**
   * The first rule of data storytelling is to focus each slide, and in fact the entire presentation, on presenting insights, not data. A lot of times, we end up using the space in the slides for presenting our analysis but not the ‘so what’.
   * For example, in the home loan case study, we can choose one of the following two options to show the information: In the first option, we are providing our audience with something that is a data finding. Some might feel this is an insight, but we do not provide them with the ‘so what’ or guidance.
2. **Know what not to present**
   * Your aim must be to present only what is absolutely critical for the audience to know at that point in time. Always remember the end objective and provide the most critical information, which will allow them to either take action or acquire information, or make decisions whatever the case may be.
   * As an example, let's recall the loan amount funnel used in the video. That is an example of descriptive data analysis.  In other words, we are looking at the data set and have found something more about its current distribution. However, it is unlikely that this is extremely critical information, except of course if the heads of the function have specifically asked for it to be included.
   * These kinds of graphs are best used for internal analyses and, if needed, included as an appendix. Including such information in the main body of the presentation would never prove to be a game-changer.
3. **Use variation** (in colour, shape, size, style) only to highlight importance and relevance, and not for appeal or aesthetics
   * Many times, we tend to use variation in colour or size without paying attention to the reasoning behind implementing such variation. As an example, take a look at the graph given below, which shows the loan amount funnel from our data set. Since there is no reasoning behind differentiating the loan amounts applied for, sanctioned, disbursed or recovered, we should use the same colour, unless we need to draw the audience’s attention to one of the bars that are more important than the others.
4. **Speak the business language**,i.e., convert data into business terminology
   * Remember that your audience is a business leader or a functional head. Therefore, write all your insights using business terminology. Oftentimes, presentations are seen to include a lot of unnecessary technical jargon, which reduces the importance or the relevance of the key message.
   * As an example, say you conducted a correlation analysis between loan amount applied for and gender. An excellent way to communicate that information would be to simply state it in layman terms, rather than to say that the coefficient of correlation between these two variables was high. Even if everyone would understand the meaning of this statement, given that it is an easy concept to grasp, it would immediately alienate the audience as you are speaking a different, non-business language.
5. **Declutter**
   * Whitespace is an underrated attribute in all presentations. As data analysts, we often look at presentations in terms of the amount of information that we should include in the slides or in terms of our communication, and less in terms of the amount of space that we leave in each slide. This is closely related to one of the previous points wherein we discussed knowing what to keep and what to leave. Whitespaces enable viewers to focus on the most important points only, thus taking their attention away from the noise.
6. **Rule of three**
   * Finally, the SME shares a very effective piece of advice on communication, which works well not just while presenting but also while writing in general: this is the rule of three. It states that “if you want something to be stuck in someone’s head, put it in the sequence of three.” It facilitates memorability and keeps the readers and the audience engaged just enough.
   * So, when you create slides, a good structure would be to keep the bullets or the key points in a series of three. As you will see in the final segment of this module, keeping not more than three bullets offers clarity and helps the audience retain the message.

And, finally, we discussed the following golden rules for preparing presentations:

1. Present insights, not data
2. Know what not to present
3. Use variation (in colour, shape, size, style) only to highlight importance and relevance, not for appeal or aesthetics
4. Speak the business language: Convert data into business terminology
5. Declutter: Use whitespace
6. Rule of three

Based on all the principles you studied in this module, let us look at the home loans case study that we shared with you during the initial phase of this module.

**Welcome to the Hospitality case study.**

Also, you learned about exploratory data analysis (EDA). Now, you will be working on a case study based on EDA.

In the case study, you will be applying all the above techniques to draw insights and help a company make decisions.

**For the case study, you will need to use:   
A Jupyter notebook to analyze data provided in CSV format,   
MySQL to read data provided in a self-contained SQL file, and   
Tableau to connect both data sets and create a dashboard.**

**Problem Statement**

**This project simulates a real-life business situation.**

**Scenario**

**You work for a European boutique hotel/home accommodation company that wants to expand to New York City, its first North American operation. The company wants to establish a network of home accommodation in New York City.**

**Your company assumes that business travelers will want the opportunity to cook for themselves rather than spend months eating restaurant food. Additionally, you expect your corporate customers to prefer home accommodations as a more cost-effective alternative to typical hotels.**

**In summary, your corporate customers and the business travelers who work for them are looking for good value – reasonable price and proximity to commercial areas of New York City. You believe home accommodation may be a money-making opportunity for your company.**

**You can read the data from the GitHub link here.**

**df\_nyc = pd.read\_csv('https://raw.githubusercontent.com/som-choudhary/Airbnb/main/Airbnb\_new.csv')**

**Background on New York City**

**New York City is massive. It has nearly 9 million people living on less than 800 square kilometers. Most of New York City comprises islands, with bridges and tunnels linking the city to itself and to the US mainland.**

**New York City comprises five boroughs, each with a different level of economic activity and number of residential units. Midtown Manhattan and Lower Manhattan are the commercial centers of New York City. Downtown Brooklyn is a distant second, although growing rapidly. Brooklyn and Queens have the largest residential populations. The Bronx is also mostly residential.**

**Each borough is made up of multiple neighborhoods (e.g., Little Italy, Park Slope, Harlem, Astoria). You can go**[**here**](https://en.wikipedia.org/wiki/Neighborhoods_in_New_York_City)**to see a list of neighborhoods by borough. But beware: Names of smaller neighborhoods can sometimes vary with data source. Just do your best.**

**The commercial neighborhoods are worker destinations, where many of the midlevel jobs are located. The commercial neighborhoods in Manhattan are Midtown Manhattan, the Financial District, and the Civic Center. The commercial neighborhoods in Brooklyn are Downtown Brooklyn: the Navy Yard and DUMBO. Many people live in the commercial areas, and many short-term rental units are also available. Many of the tourist attractions that draw people to New York City are located in or near its commercial areas.**

**Your Assignment**

**You are looking for areas underserved by your client's competitors where your client might wish to begin expanding operations. Given below is a list of questions that are intermediate facts you might want to discover and ultimately lead you to areas you might want to recommend.**

**Questions to answer**

1. **Which boroughs and neighborhoods have the greatest residential housing stock used for home-rental accommodation?**
2. **How do prices vary among the boroughs? Which borough is the most expensive?**
3. **Which room type is the most common? Which room type will appeal to a traveler whose accommodation is paid for by their company?**
4. **Which borough/neighborhood generates the highest revenue as well as the highest revenue per unit (one unit is one listing)?**
5. **How does availability differ among the boroughs?**
6. **Which boroughs should the company explore for expansion into the residential accommodation market in New York City? Why?**
7. **Which hosts have more than one property up for rent? What is the maximum number of properties owned by one host?**
8. **How expensive is real estate in NYC? How does the price of a residential unit vary by neighbourhood?**

**Imp: Task of the Assignment**

**Note: as part of the assignment, you are requested to create a dashboard of your choice answering at least 5 questions from the above. Later you also suppose to create a presentation answering the 5 questions and using key insights and dashboarding skills to explain the insights in the presentation. You will need to create a code that is explained as part of the next optional session.**

**Data Dictionary The data provided has all the listings on Airbnb in New York City.**

| **Attribute** | **Defination** |
| --- | --- |
| **id** | **Unique ID for a residential listing** |
| **name** | **A descriptive name for a listing** |
| **host\_id** | **A unique ID for the owner/host of a residence** |
| **host\_name** | **Name of the owner/host of a residence** |
| **borough** | **One of the five boroughs in New York City** |
| **neighborhood** | **A geographic subdivision of the boroughs** |
| **latitude** | **One of the coordinates for a residence** |
| **longitude** | **The other coordinate for the residence** |
| **room\_type** | **A type of residence** |
| **price** | **Price per night  in US dollars** |
| **minimum\_nights** | **Minimum number of nights for which one must rent a residence** |
| **last\_review** | **Latest review for a residence** |
| **reviews\_per\_month** | **Average number of customer reviews per month** |
| **availability\_365** | **Number of days a property is available in a year** |

**You now have the information required to start investigating the data and come up with recommendations for investments.**

----------------------------------------------------------------------------------------------------

**A Generalized Solution Approach**

All the problems that you encounter as a consultant will have some common steps to approach.

These are the points

Understand what is being asked

1. Determine the resources that you need and arrange for them preemptively
2. Work on subtasks simultaneously
3. Initial steps in the solution-building process:
   1. Download required software
   2. Read and access data
   3. Set up a ReadMe file
   4. Get a feel of the data

Once you have completed these steps, you must proceed to cleaning the data. The missing value problem is the most challenging one to solve in the data-cleaning exercise

Before conducting data analysis, you will have to think about other considerations, such as data transformations.

This is what you learned about in the video

1. Data transformations
   1. Unit conversions
   2. Encoding
   3. Variable calculations
2. Data exploration
3. Checkpoints

After you have done all this, the last step will be to draw a conclusion(s) and present your findings. In the video below, Sean will explain how you should present your findings and conclusion(s).

----------------------------------------------------------------------------

**The below is the link for new York city and the same I have copied below the url**

<https://en.wikipedia.org/wiki/Neighborhoods_in_New_York_City>

the link to data is

<https://raw.githubusercontent.com/som-choudhary/Airbnb/main/Airbnb_new.csv>

​