

Data Storytelling

Introduction to Data Storytelling

The term storytelling is used widely, although data storytelling hasn't quite become popular. Data storytelling is, essentially, a practice that is developed by analysts and data scientists over the years to communicate their findings and insights from data to technical and non-technical business stakeholders.

To solve a business problem, you need to first understand it. Then, you need to identify the data set or gather data, followed by working on data preparation and cleaning methods, and then building models to derive key insights from the data. Finally, you work on model evaluation. After deriving key insights from the data, you would often find yourself in situations wherein you would have to use some medium, for example, for the purpose of visualisation, to explain these findings or reports to a technical or a non-technical audience, or a senior business audience. In such a case, it is important to understand the important elements that you need to include while communicating the key findings from the data.

This is not as easy as it sounds. We spend so much time on the initial steps of project planning and evaluation that we often ignore the purpose of the analysis, which is to either get more information on any particular topic, or help the senior business management take important decisions based on our findings.

Here, you will:

- Learn what data storytelling is
- Learn about the importance of data storytelling
- Understand the important elements of a good story with data:
 - Narrative
 - Structure and Flow
 - Visuals
- Learn about the five principles of insights, the Pyramid principle, storyboarding principles, etc.
- Learn about some of the best practices related to data storytelling

Pre-read Assignment

This topic is based on effective communication and presenting your data findings to business and technical stakeholders. To understand the concepts well, let's take a look at a problem statement and the data set for that problem (the 'Home Loan data set', which is provided in the module), and go through some of the insights that we have derived from this data set. After performing a simple analysis of the data set by using simple pivot tables, you will understand how we derive these insights from the data set. Your task is to look at the following insights, use pivot tables to cross-check these insights from the data set, and, while doing that, use appropriate graphs, which would help you present these insights.

Problem Statement: Consider you are working for a bank and are the lead analyst working on the Business Analytics team. You need to analyse the Home Loan data set, which contains data related to customers, home loans, product channels, etc. You have been asked to present your findings in a few weeks to the following two business unit heads: Head of Sales and Head of Operations. The Head of Sales is responsible for increasing the sales of loans and associated products, while the Head of Operations is responsible for managing branches, channels and other associated operations.

The bank is witnessing growth in sales year-on-year but at a declining rate. The Head of Sales believes that business conditions are responsible for this and that there is nothing much that the bank can do. On the other hand, the Head of Operations has had a budget cut and so, very less effort has been invested into making the distribution channel effective.

Insight Generation: Your job is to analyse the data using basic data exploration and visualisation techniques and derive insights related to the points given below. Remember, this exercise is less about the insights themselves and more about the way you communicate them. Therefore, focus less on the data analysis and exploration techniques, and more on how you would put them as a story in a Presentation. You can create all of your graphs using Excel and copy-paste them to PowerPoint as images. We will provide you with certain insights, but you can feel free to explore more insights based on the questions outlined below:

1. What does the customer trend indicate?
2. What can you tell us about the customers, their demographics and their purchase behaviour?

3. What are the top three findings that you would like to share about the trends in home loans?

To help with your data exploration exercise, here are some insights for you:

Insights on Loan Amounts

- 87% of loan amounts applied for have been sanctioned
- 92% of loan amounts sanctioned have been disbursed
- 80% of loan amounts applied for have been disbursed
- 53% of disbursed loans have been recovered

Insights on Trends in Loans Applied For

- In Q3 2019, the loan amount applied for was the highest in the last 3 years (since January 2017), at ₹8,111 lacs.
- The lowest amount during this period was ₹5,920 lacs in January 2017.

Insights on Customer Channel Preferences

- Between 35% and 36% of customers prefer online channels, irrespective of gender.
- Nevertheless, younger customers in the age bracket of 26–33 years tend to use online channels more, with about 50% preferring to use online channels and none of them preferring to use tie-ups.

While adding more insights, remember that the goal is to understand the elements that you would include in your presentation and how you would structure it. You will have to present these insights with just a few slides and assume that you are presenting to the senior-most sales executives.

What is Data Storytelling?

Let's take an example to understand what data storytelling is: A cholera outbreak in London claimed the lives of many people. This was not the first or the second but the third outbreak in the last 20 years. Jon Snow, a famous physician, tried to communicate his findings and hypothesis to several people but was unable to convince them. He used qualitative evidence to convince individuals and the government that the disease had spread through contaminated water. Unfortunately, his interview notes, research papers and written summaries did not work.

Now, let's go over the story once again.

In 1954, a cholera outbreak in southern London claimed the lives of over 600 people – 500 in the first 10 days itself. Interestingly, this was not the first or the second time that this had happened. During the 23 years spanning 1831 to 1854, the city had witnessed 3 major outbreaks of cholera, claiming the lives of over 32,000 residents of London.

At this time, a physicist practising in London, John Snow, had the opportunity to study this disease not once or twice, but three different times.

In order to convince the people about his hypothesis, he conducted extensive research backed with data, strong visualisation and an excellent narrative. He was able to uncover patterns that would otherwise be impossible to derive. He plotted the data on a map, showing that a water pump in Broadway was the source of the infection. Here is a snippet of the data plotted on the map.



With this method, he was able to inform the authorities about the way the disease had spread. This changed the mindsets of doctors and hospitals, and London's infrastructure improved drastically as a result of his finding.

So, quantitative analysis and a data-backed story helped John Snow prove his hypothesis. He combined scientific methods with statistical analysis and has been credited with one of the first recorded investigations in the field of epidemiology. With his approach, he could help save many lives and it brought about a transformation to the infrastructure of cities and towns, in addition to the introduction of several new approaches for dealing with infectious diseases.

This is the difference between a simple story and a data-backed story.

When certain facts are presented with solid data and strong visualisations, they have the power to change minds. Any story founded on solid data analysis, strong and clear insights, powerful visualisations and straightforward communication techniques has a much higher likelihood of achieving its goals – whether to inspire, persuade or inform – than non-data-backed stories.

Importance of Data Storytelling

In today's world, wherein terms such as big data, insights and analytical techniques, are receiving increasing emphasis, one might wonder why the topic of storytelling has all of a sudden become so important; why have terms such as 'data journalism', 'data-based stories' and 'data-based insights' been coined; and why have experts begun viewing the communication of insights as one of the most critical skills in the field of data analytics.

There are several reasons for this.

Access to advanced tools and technology:

- With improved access to analytics tools and capabilities, as well as availability of analytical talent, we can access and analyse data easily and much faster, and derive insights. The reasons behind this are the availability of easy-to-use-tools, abundant self-service analytical capabilities and free software, which allow analysis of vast amounts of data in a short period of time.
- Not only do these tools allow us to easily explore large amounts of data, they also save us a lot of time. You can run medium- to high-complexity analyses in a matter of hours without any formal degree in programming or coding.
- As an example, you can use Excel, PowerBI or Tableau to easily analyse data on the sales of a retail store for an entire region, you will easily be able to analyse data using excel, PowerBI or Tableau. The problem that you would face will not at all be related to the availability of data or analysis. The core problem would be to understand how to communicate the huge amount of data points and insights easily and clearly to the Head of Retail Strategy or the Head of Channels.
- As we can see, due to the huge abundance of data and associated advances in tools and technology, decision makers face no challenges in arriving at insights. The problem lies in the lack of time with your audience. It is now more important than ever for senior leaders to not just have a great analysis, but also have someone to communicate it effectively.

Too little time:

- Unfortunately, a huge abundance of data does not necessarily mean that we have a lot of time to analyse all of it and derive findings. This, in turn, means that all of the data and the insights are rendered meaningless as senior executives, who need to understand the data, are often short on time. Consequently, they need someone to help remove noise and communicate only what is important.
- Going back to the retail store example, if you are the Head of Channels and are asked to make a decision on which stores to keep open and which ones to close during COVID-19, then you will not have any shortage of information on which to base that decision. Within a few hours, you will have all the raw data that you need right from store sales to regional information, walk-ins, top-selling products, pricing and trends.
- It is also quite likely that you will have access to a team of one or two analysts who will help you derive hundreds of data-based findings. What you will struggle with, however, is how to quickly understand all of those findings and make decisions, and how to communicate all of your findings to your seniors, leaders, colleagues and other functions?

Considering the broad factors mentioned above, data storytelling becomes quite an important task, and it comes in handy to bridge the gap between junior and senior stakeholders, as well as the technical and the non-technical business stakeholders.

Characteristics of a Good Story with Data

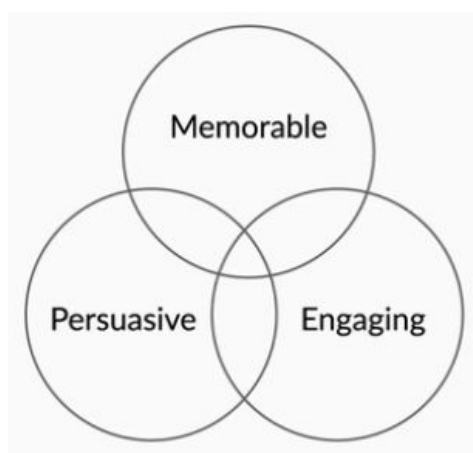
In today's business context, data storytelling becomes critical due to the following three characteristics: Memorable, Persuasive and Engaging.

Memorable: As per a Forbes article on why data storytelling is essential, the use of data in our presentations allows us to help the audience remember key messages. For example, in a study conducted by Chip Heath, the author of *Made to Stick* and a Stanford professor, it was found that over 63% of the participants were able to remember stories, but only about 5% of them were able to remember a single statistic.

Persuasive: Similarly, another study quoted by the same article compared the per-participant donation amount generated from a campaign by the Save the Children foundation. The research showed that the per-participant donation amount generated through a story-based version of the campaign was more than double the amount generated from an infographic version of the same campaign.

Engaging: Finally, while sharing our findings, if we present the audience with a collection of numbers or graphs, and if they have to do the work of either going through the content or

making sense of it, then they would either become confused and stop paying attention, or may become more interested in nitpicking the details and the methodology, instead of understanding the core message. So, if your story is not engaging, then you quickly lose the attention of your audience.



In what comes up next, you will understand that despite knowledge of all the key elements of data storytelling, people often struggle to communicate their findings effectively. This makes it important for us to know what makes data storytelling such a difficult task.

What Makes Data Storytelling Difficult?

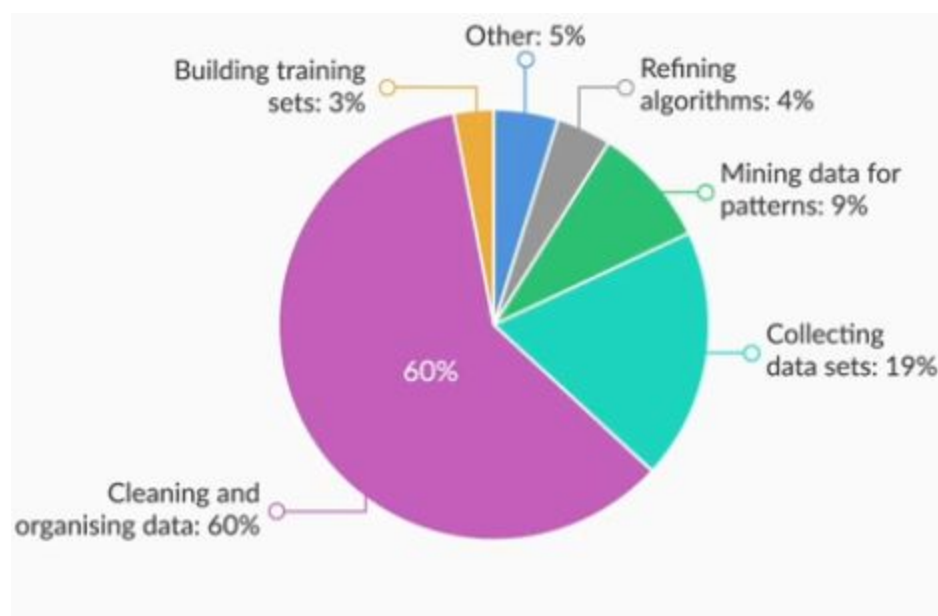
So far, you have learnt about the various factors that make data storytelling important when closing a data science project. But why is it so commonplace to see people struggle to communicate and present their findings, despite knowledge of the key elements of data storytelling? Why do many leaders continue to make ineffective decisions due to lack of understanding of insights even after investing money and time on data science projects?

The reason is that they often spend disproportionately huge amounts of time and effort in performing each of the following tasks:

1. **Data collection:** This includes gathering data through various means, such as surveys, customer interviews, vendor interactions, online metrics.
2. **Data preparation:** This includes updating, cleaning, trimming, identifying and eliminating outliers.
3. **Data wrangling:** This includes building variables, categorising variables, etc.

4. **Data analysis:** This includes the use of various tools and methods to visualise data, explore its distribution and conduct basic to advanced analyses.
5. **Data insight generation:** This is the step wherein you put yourself in the shoes of the end consumer of the analysis and arrive at the final insights.
6. **Communicating data findings:** And finally, this is the step wherein you communicate your data findings to the audience.

The pie chart shown below depicts the general proportion of time that is spent in the tasks described above.



Of all the steps mentioned above, the first four are the ones where we spend the most time. The lower the tenure of the project, the higher the proportion of our time that goes into preparing the data.

In fact, as per some estimates, data preparation and cleaning can take upwards of 60% of a data analyst's time. In fact, a study conducted by Forbes showed that data preparation accounts for over 80% of a data scientist's time. They spend about 20% of their time collecting data, and then 60% on cleaning and organising data, altogether spending around 80% of their time on preparing and managing data for analysis.

Considering most of our time, as analysts, is spent on these activities, we fall in the trap of trying to justify this time in our final presentations. For example, if a team lead is asked to show the progress on the analysis of a complex data set, say, the data of online traffic on a web portal, then they would try to show everything that they did during the 80% time when they were

preparing and managing data for analysis, instead of showing what they did during the 20% time wherein they actually derived the insights.

So, naturally, when we begin to communicate our results, insights or project progress, we like to talk about the various steps that took the most time. For example, how we gathered data, how we prepped it, what were challenges that we faced and what was our methodology. In fact, we also like to talk about the complex approaches that we adopted, in order to justify the accuracy of our results and quality check procedures. While this is interesting to us, it is less likely to be interesting to others.

This happens when we do not understand either the objectives of the meeting or the background of our audience. When this happens, we end up getting confused about how much detail is needed and how much is too much. Since we do not know our audience that well, we over-prepare for a more technical audience and less for others.

So, what does one do in such a situation? How do we ensure our analysis is not just understood clearly, but also has a lasting impact on the audience? How do we make sure the audience remembers what they hear and are able to take action based on it? How do we make sure we do not just inform the audience, but also influence the outcomes?

One of the ways to convert a dry, dull and extremely elaborate presentation to an informative yet engaging one is to use the technique of storytelling. Simply put, this means communicating the results in a manner that is interesting, impactful and memorable. As we saw above, these are the attributes that drive change. Without powerful communication, your analysis becomes meaningless and a wasted effort, whereas the right communication style can make your insights impactful and help you achieve your end objective.

Session 2: Components of a Good Story with Data -

I

Essential Components of a Good Story with Data

Now that you have learnt what data storytelling is and also understood its importance in today's business world, we will now learn about the essential components that make a good story. Let's begin our learning with an example.

In February 2018, ABC's business analytics team uncovered the reason for the company's continued decline in sales. They gathered data from the past 5 years and conducted detailed analysis, including correlating sales with various other expenses, such as marketing, advertising and technology spend. The sales head of the company had to present to the CEO in a few days and it was becoming increasingly challenging for him to filter what was important and to explain what was going on. So, he decided to work with the business analytics team to conduct further analysis. Sales data related to pipeline indicated that the root cause of the problem was not related to pipeline conversion but, in fact, to the quality of leads. The business analytics team gathered data related to sales by product line and geography, and used PowerBI to map the variables to the external market as well. Interactive dashboards showed that there was a blip in sales in Q3 of 2012 and Q1 of 2013. ...

If you observe the story above well, then you will notice that it is not very narrative in nature or visually appealing, and also does not have a proper flow and structure.

Given below are a few elements that seem to be missing in the story. We will briefly explain the elements now, but will take a look at these in detail with examples later:

1. **Narrative:** A narrative is the core of a good data story. It means connecting the dots, or connecting the various aspects together. It provides a summary or an explanation of what the data tells, why that is the case and what to do about it. A presentation that contains no narrative is one wherein the client or the audience is presented with a random set of slides, with no single narrative tying them together.

Key questions to answer while addressing this aspect: You need to ask certain key questions with regard to the aspect of Narrative. These include the following: To whom are you presenting and why? Why should they bother and why would they listen to you? How will you simplify the takeaways? How will you connect your insights and make them 'stick'?

In the example shown above, there is a collection of facts but there is no story. It is unclear who we are talking to and what we are trying to say. Simply put, the paragraph is missing a narrative.

2. **Structure and Flow:** Structure is the next aspect of a story. It is a roadmap, which ensures a crisp and logical flow of the presentation. It outlines the journey that your audience will take with you in those 30, 45 or 60 minutes.

Key questions to answer when thinking about this structure: You need to ask certain key questions with regard to the aspect of Structure. These include the following: What are the sections that you will cover and when? What does your audience need to

know? How much content will you cover? What comes first and what comes next? How will you simplify the flow of the presentation?

Again, going back to the example, the story does not have a logical flow. At times, it talks about the approach that they adopted and the steps that they took, while at other times, it talks about the challenges that they faced. A story with good insights but a bad structure would leave the audience confused and cause them to become inattentive to your presentation.

1. **Visuals:** The final aspect of a story is visuals. This includes fonts, colours, graphs, diagrams, interactions, etc. By including effective visuals in presentations and working through important design principles, we can make our story interesting and engaging.

Key questions to answer when thinking about visuals: There are certain key questions that you need to ask with regard to the aspect of Visuals. These include the following: How will the audience absorb the information? How will you present all of the information and content on one page? How will you simplify the graphics? How will you make your slides visually appealing and engaging?

Importance of Objective and Agenda

Before beginning a presentation, you must have a clear objective for the presentation itself. You can define the objective of the presentation by:

Starting with certain ‘why’ questions: You must try to understand why you are presenting. What is the key reason that you have been asked to present? What is the audience expecting to get from the time they invest? There are two main reasons why we need to present: we need to either provide information or undertake action.

Following are points related to information (FYI):

- To provide information
- To explain the process
- To share challenges

Following are the points related to action (FYA):

- To make a decision
- To influence an outcome
- To get clarity on next steps

Information should be provided for the following two reasons: simply to provide information and to take action on certain projects.

FYI is important to technical stakeholders or business leaders who are not very senior stakeholders in a company but are, at the same time, interested in knowing the key technical aspects of a project. For these individuals, information is provided to:

- To simply present information,
- To explain a process and
- To share challenges.

FYA is important to senior management or business leaders who do not want to get into the details of the data preparation or the data evaluation process; rather, they want to understand the key insights, which further aid the decision-making process. For these individuals, information is provided to:

- To help them make a decision,
- To influence an outcome and
- To get clarity on the next steps.

As an example, if our objective is to help the audience take a decision, then we need to understand whether we need to provide them with more data or more information, or more clarification? Is the meeting about convincing the audience or about reaching a consensus? Identifying 'the one thing' or the 'elevator pitch'. If there was just one sentence that you could communicate to the team, what would that sentence be?

When we are clear about the objective, it is easy to identify the content that needs to be included in the narrative and also determine its structure. It also helps us prepare more, and include what is important and leave out what is not as important.

Note that the audience determines the objective of your presentation and it is never the other way around. So, if your audience includes mid-level executives, then it is likely that they will be more interested in the information and the approach, since they will either present it further, or it is their job to ensure the analysis is robust and the findings are accurate.

Similarly, if your audience includes senior executives, then the information that you include in your presentation should be action-oriented and the key insights should be able to tell the business leaders what action needs to be taken thereafter.

Now, let's go back to our example of the home loan analysis. Imagine you are the lead analyst on this project, and you have been asked to deliver a presentation. The narrative, visuals and the structure of your presentation will change depending upon its objective.

If the audience includes mid-level data function executives and not the leaders of the business, then the objective of the presentation would be to provide information. As a result, most of your content would be about what information is needed and why. They will be interested in knowing about the approach that you adopted, the methodology that you chose, the types of data sources that you have used, the assumptions that you have made and the challenges that you are facing while determining the next course of action.

However, if the audience is a business head, say, the head of sales, then they will not quite concern themselves with any of the information above. In fact, in this case, you must include nothing but key findings, i.e., why is there a decline in the growth of home loan sales and what to do about it, or how are your customers' preferences changing and what approach should the sales teams adopt in order to change.

Narrative - I

The first part of a narrative relates to knowing more about your audience. There are multiple ways to view your audience, but the following two factors are critical to understanding whether you can manage to know about them: first, to understand their backgrounds and roles and, second, to know about their seniority. While this is not always possible, you should prepare as much as you can.

Now, let's take a look at the following matrix. The x-axis shows people who are from a business background, a technical background, as well as those who have both a business and a technical background, in an organisation. On the y-axis, we have people with a lower tenure and a senior tenure, as well as those who have both.

BOTH	<ul style="list-style-type: none"> Solve for both. Provide everything that the senior most needs and put the rest (junior/mid-level) items in the appendix or as separate slides 	
HIGH TENURE (Senior)	<p>E.g.: C-level or VP or Senior Directors of business functions</p> <p>What's in it for me?</p> <ul style="list-style-type: none"> Show me metrics that I care about Help improve my BU's performance so use business language Show me what you found not how you found it Don't show me data prep, methodology and analysis Don't use technical jargons in the presentation 	<p>E.g.: C-level or VP or Senior Directors of technical functions or data leaders or</p> <p>What's in it for me?</p> <ul style="list-style-type: none"> Show me what you found <u>and</u> how you found it I need to understand data prep, methodology and analysis BUT not during the presentation. Either before or after and ONLY when I ask for it. Don't use technical jargons in the presentation (yes, even though I understand it)
LOW TENURE (Up to mid-level)	<p>E.g.: Managers / Senior Managers and team leaders within business functions</p> <p>What's in it for me?</p> <ul style="list-style-type: none"> Show me metrics that my boss cares about Help improve my BU's performance Show me what you found and how you found it but only in a very simple language so I can explain if needed Provide me confidence that data prep, methodology and analysis is robust Don't use technical jargons in the presentation that either I or my boss won't understand 	<p>E.g.: Managers / Senior Managers and team leaders within technical or data functions</p> <p>What's in it for me?</p> <ul style="list-style-type: none"> Show me what you found <u>and</u> how you found it I need to understand in a very detailed way data prep, methodology and analysis AND put them in the main body of the presentation. If needed send me analysis in advance and set up a call to explain. Feel free to use technical jargons in the presentation (yes, even though I understand it)
X	BUSINESS BACKGROUND / NON-TECHNICAL	TECHNICAL BACKGROUND / NON-BUSINESS

For people with a **business background** and a **low tenure**, which could include **Managers/Senior Managers** and **team leaders** within **business functions**, the following points must be considered while presenting:

- Show the metrics that matter to your bosses
- Help improve the performance of your business unit
- Show them your findings and how you derived them, but only in very simple language so you can explain if needed
- Give them the confidence that the data preparation, methodology and analysis are robust
- Do not use technical jargon that either you or your boss will not understand

For people with a **technical/non-business background** and a **low tenure**, which could include **Managers/Senior Managers** and **team leaders** within **technical/data functions**, the following points must be considered while presenting:

- Show them your findings and how you derived them

- You need to have a very detailed understanding of your findings, including the data preparation, methodology and analysis, and include these in the main body of the presentation. If needed, you can send them an analysis in advance and set up a call to explain
- You can use technical jargon in the presentation

For people with a **business background** and a **higher tenure**, which could include **C-Level/VP/Senior Directors of business functions**, the following points must be considered while presenting:

- Show them the metrics that matter to you
- Help improve the performance of your business unit and, therefore, use business language
- Present to them your findings but do not show how you derived them
- Do not show them the data preparation, methodology and analysis
- Do not use technical jargon in the presentation

For people with a **technical/non-business background** and a **higher tenure**, which could include **C-Level/VP/Senior Directors of technical functions** or **data leaders**, the following points must be considered while presenting:

- Show them your findings and how you derived them
- You need to understand the data preparation, methodology and analysis but not detail them out during your presentation
- You can show them these details when you are asked to
- Do not use technical jargon in the presentation

Five Patterns of Insights

Insights are yet another important element of a good narrative in a story. Now, you might have noticed that there is a certain underlying pattern to the insights that we provided to help with your data exploration exercise. What this means is that irrespective of the business problem that any industry is trying to solve, the insight that is generated has some common patterns.

As explained above, any business insight can be said to have the following five patterns:

1. **Unknown Result:** The number is unknown and is of significance:
 - a. Zimbabwe has a population of this much.
 - b. Product return is 18%.

2. **Surprising Extreme:** Some value are larger and some are smaller, which is not the expected result:
 - a. AAP has the highest number of criminal cases against them
3. **Surprising Comparison:** When you compare data and the findings are unusual and surprising
 - a. Error rate is 5% instead of 1%.
 - b. Lucknow's mortality rate is five times higher than the national mortality rate.
 - c. Bengaluru is almost 5 degrees cooler than Hyderabad.
4. **Significant Outliers:** When you have an extreme of an unusual form, far more than you expected:
 - a. Nandan Nilekani is richer than the rest of the Congress party put together.
 - b. The call-drop rate in Mumbai is five times higher than that in any nearby city.
5. **Abnormal Distribution:** when you see abnormal distributions or sudden peaks in some value
 - a. The number of students scoring 50, 60, 70, etc. is much higher than the number of students scoring 49, 59, 69, etc.
 - b. Temperature in Bilaspur in Chattisgarh is high when it is low in the surrounding areas, and vice versa.

What you just learnt is a robust classification with which you can categorise all the insights that you have derived. This framework creates grounds for you to ask questions, although you also need to know the area well to ask better questions.

For example, if you have the sales revenue data for a list of products manufactured by a company, then you can ask specific questions beforehand, such as:

- Is there a product that is showing an abnormal distribution in its sales revenue trend?
- Is there a product that is selling unusually better than other products?
- Is there a product that sold far better than the expected historical standards?

In this way, you can tailor your analysis effectively to be on the right track while deriving insights from data.

At this point, it might be natural for you to ask the following question: “I know what I need and where to look for insights, but how exactly should I look?” This translates to the following question: “What are the procedures that I should apply on the data set so that I can generate insights?”. You will get to learn about this in the next two segments.

Analysis Approach: Deriving New Columns

In the previous segment, you learnt about the five simple patterns through which you can classify any insight. In this segment, you will learn about the methods with which you can actually generate these insights.

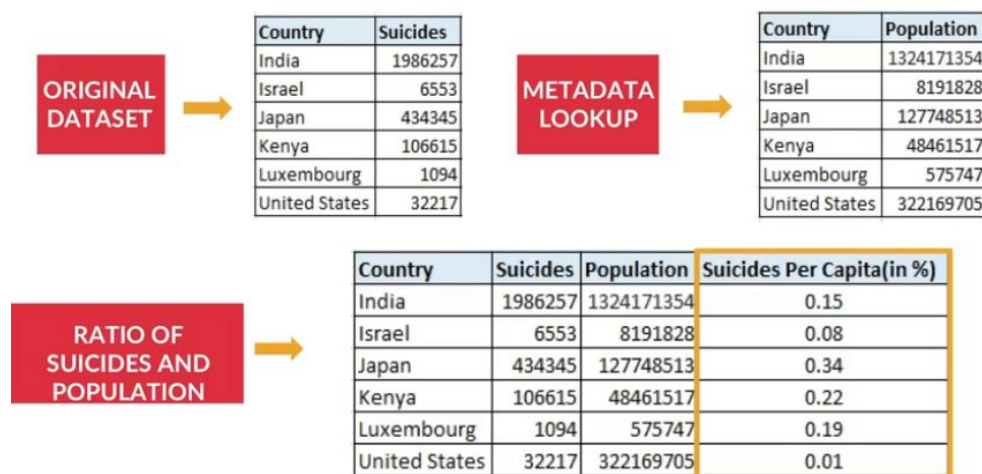
Although the five patterns discussed previously are sufficient to categorise insights, the data set that you have may not necessarily be in a format on which you would apply these patterns directly. Therefore, you also need to learn a few techniques to analyse your data so you can apply the five patterns and start generating insights. These techniques can be categorised into the following two types: exploratory data analysis and hypothesis-driven analysis. These analysis techniques utilise the following two methods of data manipulation to derive insights: creating new columns and reducing the number of rows.

Let's start with the first data analysis technique with which you can analyse data to derive insights: creating new columns.

Let's take an example to understand this technique better. Consider you have a data set, which only contains revenue and cost information. You can go ahead and create a separate column where you would be calculating the profit and applying the five patterns. Some of the most common methods of adding new columns to a data set include the following:

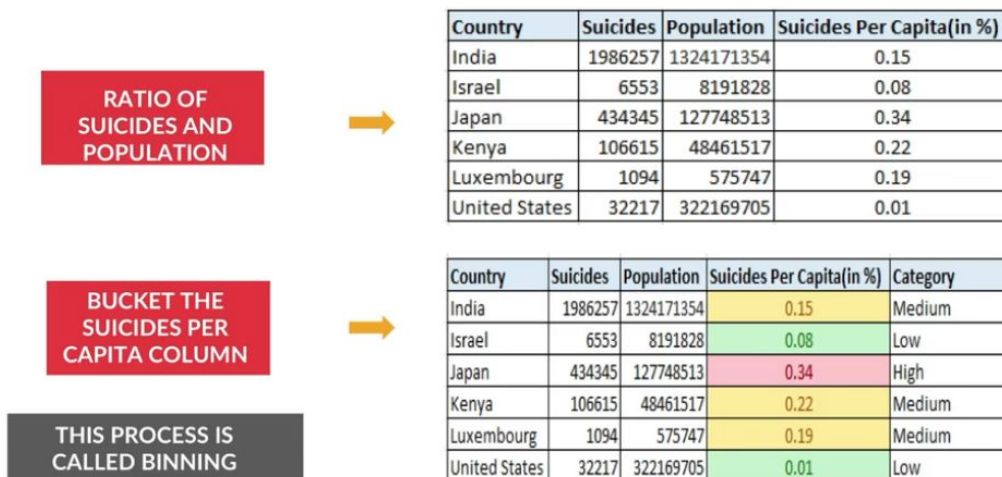
- **Metadata Lookup:** Metadata is essentially an additional data set or a data sheet that is available to you and which provides information on the original data. As an example, you would get the following result if you had data on suicides in various countries and the population of each country was available to you on a separate sheet, and you performed a VLOOKUP to create a new column in the original data set.

DERIVING NEW COLUMNS



- **Calculations:** You can perform a variety of calculations using the numeric columns in your data set. For the example above, you created a new column named 'Suicide Rate (in %)' from the 'Suicides' and the 'Population' column.
- **Binning:** This process essentially bins a given numeric column to specific categories. In the example above, you converted 'Suicide Rate' into specific bins and categorised them as High, Medium or Low.

DERIVING NEW COLUMNS



- **Business-Specific Metrics:** This part would be specific to your domain and, hence, the metrics or the [KPIs](#) that you might be using would be a useful additional column.

These methods may not seem exhaustive, as a variety of analysis procedures can be used to derive new columns. Nevertheless, they more or less encompass a standard approach that you should adopt once you have the data with you to check for new insights.

Another nifty way of creating new columns, and one which is used heavily right now, is through machine learning.

The various machine learning techniques that you can use to create new columns are as follows:

- **Classification**
 - Classify loans as good or bad
- **Clustering**
 - Cluster customers on the basis of their behaviour to specific groups
- **Time-Series Analysis**
 - If you have the pricing data for the past 10 years for several products, then you can use forecasting algorithms to predict future prices.
- **Feature Extraction**
 - Sentiment analysis on Twitter
 - Competitor analysis
 - Quality check
- **Sentiment Analysis**
 - Sentiment analysis on Twitter

To summarise, there are mainly two techniques with which you can create new columns in a data set: by performing calculations and through models - either statistical or machine learning. Statistical models often take a sample of the original data and use it to infer the behaviour of an entire population, whereas for machine learning models, we run algorithms on a set of predefined data, called 'train data', to build a model and then apply the model again on another set of data, called 'test data', to test its accuracy and precision. If you find it difficult to understand some of the terms in the previous sentence, don't worry. You will learn about all of these concepts in detail in the next two courses. For the time being, a cursory understanding of the difference is sufficient.

After deriving new columns, you can go ahead and apply the five patterns that you learnt about earlier to generate insights.

Analysis Approach: Summarising Rows

Now that you have learnt about the various methods with which you can create new columns to apply the five patterns in order to derive insights, you will next learn about another method of summarising the information provided by a data set: by collapsing rows.

The following procedures are mainly used to summarise rows to analyse data for deriving insights:

- **Summarising Numeric Columns**
 - Add all the rows
 - Find the average
 - Find the median
 - Find the mode
- **Grouping on the Basis of a Column**
 - In the snippet given below, we are collapsing row on the basis of states.

Product Name	District	State	Total Sales
P1	D1	S1	1000
P2	D2	S2	1300
P2	D3	S1	1200
P1	D1	S3	1400
P1	D2	S1	1150
P1	D3	S2	1210
P2	D3	S3	1100

State	Total Sales
S1	3350
S2	2510
S3	2500
Grand Total	8360

- **Pivot table:** Collapse at state level and compare.
 - In the example given below, we can check whether there are any products that are strongly associated with any state:

Product Name	District	State	Total Sales
P1	D1	S1	1000
P2	D2	S2	1300
P2	D3	S1	1200
P1	D1	S3	1400
P1	D2	S1	1150
P1	D3	S2	1210
P2	D3	S3	1100

Sum of Total Sales	Product Name		
State	P1	P2	Grand Total
S1	2150	1200	3350
S2	1210	1300	2510
S3	1400	1100	2500
Grand Total	4760	3600	8360

- Machine Learning

- Predict the sales of one product on the basis of the sales of three other products and calculate their importance
 - The coefficient of the predictive model can become a new data set.
- Using relationships between columns, for example, correlation

BANK SALES DATA (NUMBER OF UNITS SOLD)			
City	Credit Card	Loans	Saving Accounts
C1	1000	1200	650
C2	1100	1350	120
C3	2100	2050	800
C4	1300	1220	1200
C5	3000	3300	2210
C6	2100	1700	3100

CORRELATION MATRIX			
	Credit Card	Loans	Saving Accounts
Credit Card	1		
Loans	0.94378061	1	
Saving Accounts	0.674049061	0.458641	1

Note: Correlation as a concept will be covered in detail in a future module. For now, it would suffice to have a basic understanding of what correlation does, i.e., it measures the strength of the relationship between two variables. The value generally ranges between -1 and 1. A higher positive correlation indicates a positive influence of the variables on each other. For example, if you find that the correlation between the sales of Product A and Product B is 0.7, then you can

say that they sell well together. You can go [here](#) to get a brief idea of how correlation works and learn how to create a correlation matrix in Excel [here](#).

Insights: Overview

Now, you will learn about two techniques with which you can generate new information from a given data set on which you can apply the five patterns to gather insights. It may not be evident right now, but these two analysis techniques are quite powerful in generating insights when you cannot proceed beyond a certain point and feel that you have performed all possible analyses.

The two generative techniques are as follows:

- Generate insights and get unstuck even after a series of analyses
- Even with a small data set, you can perform hundreds of analyses

At this point, you may have a question in mind: “I understand that I’ll be able to derive hundreds of insights from the data by repeatedly following the procedure given above. But how do I know which insights are useful for me?” In others, it means to be asking how to know which insights should be given more importance than others.

To do this, you can make use of a certain checklist to go through the entire data set, perform a variety of analyses – by either collapsing rows or preparing new columns – and infer whether or not the generated insight is useful based on the results.

The qualifying criteria for insights include the following:

- Is it useful?
 - Is it related to the problem that you are trying to solve?
 - Is it actionable?
- Is it interesting?
 - Is it non-obvious?
 - Is it unusual?

Also, it is always a good idea to have a list of hypotheses ready when you are preparing to analyse certain data. You can formulate them either on the basis of the five patterns that you learnt previously, or you can utilise the business aspect of the problem at hand. Combining both of them, you can generate insights and see if it leads to something useful and interesting.

Now, if we take a look at the home loan example, then we will see that we have derived plenty of insights based on these patterns. For example:

1. The trend in home loan amount shows a surprising increase in amounts sanctioned in July 2019.
2. Similarly, if the industry average of recovered loan amount ratio is 50%, but it is only 35% in our company, then it is a surprising insight that we can include in our summary.
3. The fact that younger customers use online channels more than older customers is not an insight because it is obvious. However, if the percentage is either lower than the average, or if there is an unexpected deviation, then that becomes a useful insight.

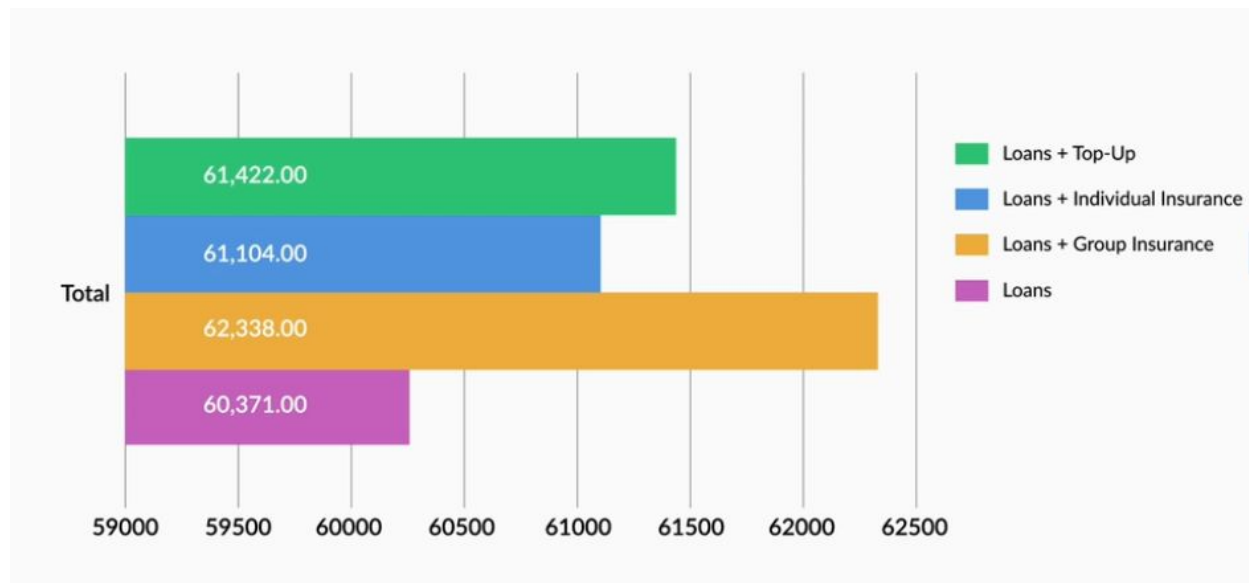
Narrative - II

Finally, after understanding the objective of your presentation, you must pay attention to the context. Context refers to showing why the results are meaningful and relevant to us. Why should someone care about all of that information? Many times, we achieve this by including information related to our own past performance, challenges and financials, profitability, industry, competition, historic figures, etc. The idea is that the audience should not be seeing the figures in isolation but in relation to some context.

So, in our home loan example, context will include information on how the market is. However, even more important is why they should care about any of the information that you are providing them.

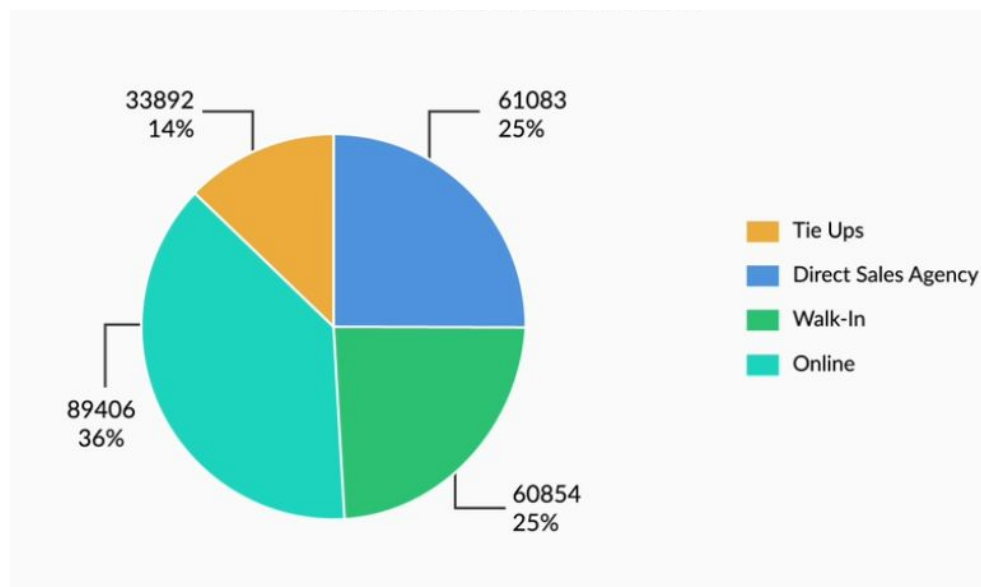
So, let's try and understand the meaning of narrative through our home loan example. We can draw multiple insights on loan amounts, top-selling products, channel distribution, etc. Now, imagine you had to present your findings to the Head of Sales and your objective was to provide information that enabled them to decide which customer segment to focus on and why they should focus on it.

Following is bar chart of 'SUM OF APPLIED FOR LOAN AMOUNT FOR DIFFERENT PRODUCT CATEGORIES':



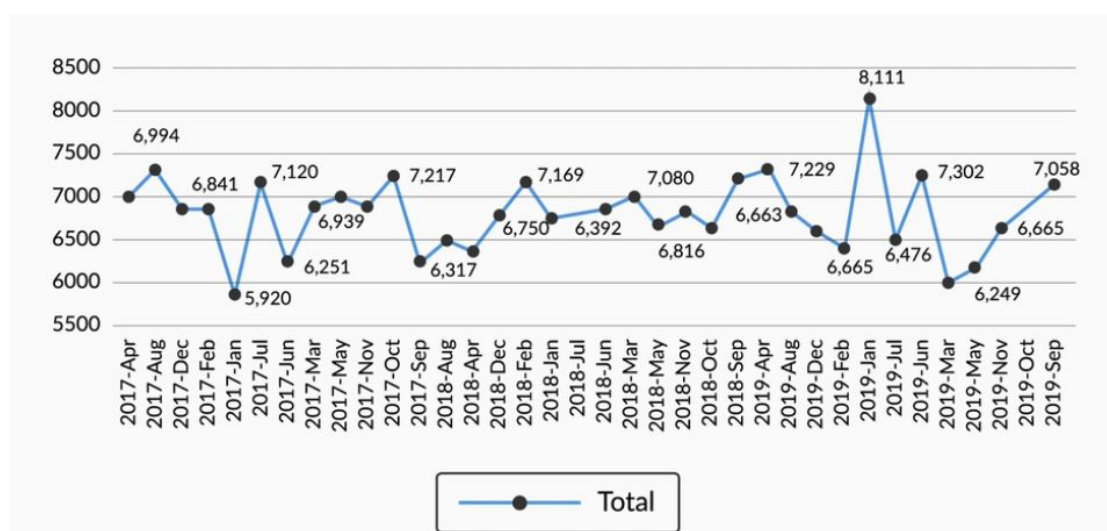
As we can see from the graph given above, the largest loan amount is under 'Group Insurance'. Now, from this graph, the Head of Sales might already understand that 'Group Insurance' has the largest loan amount; so, we can just tell them just how large the amount is. However, to the Head of Operation, this will probably not be meaningful unless we dig a little deeper and tell them, for example, that the 'Loans + Group Insurance' category has the highest number of applications but with the lowest sanction rate. So, the information that we provide can be an insight for one scenario and fruitless in another.

Here is a pie chart for channel distribution:



Now, in the graph given above, online is the most popular channel for applying for loans. This information will be extremely useful for the Head of Sales, Head of Distribution and Head of Operation. We can make the graph more meaningful by adding comparisons or showing trends. We can add a context by showing how other banks are tracking or how online channels are behaving in general across financial products. This has absolutely no context to a client unless they ask for it.

Here is the distribution of 'LOAN AMOUNT APPLIED FOR':

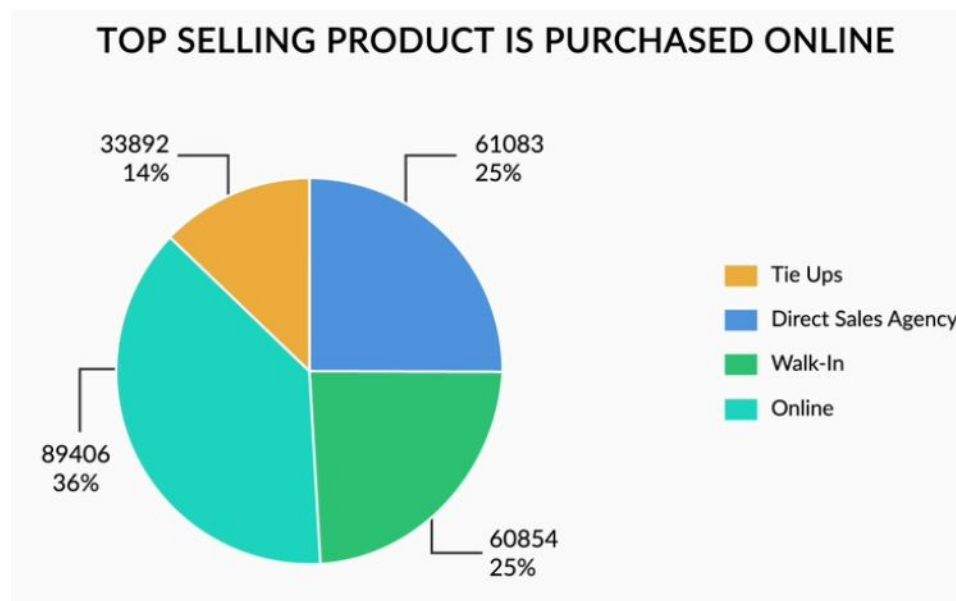


Now, in the graph given above, we can see that in July 2019, the applied loan amount was higher than that in the previous quarter. It is good information but not actionable.

These graphs, although accurate, do not provide a lot of actionable information to the audience. In addition, they are hard to remember and are not personal, i.e., they are not ‘talking’ to anyone. Consequently, the narrative has become weak and some key insights are missing.

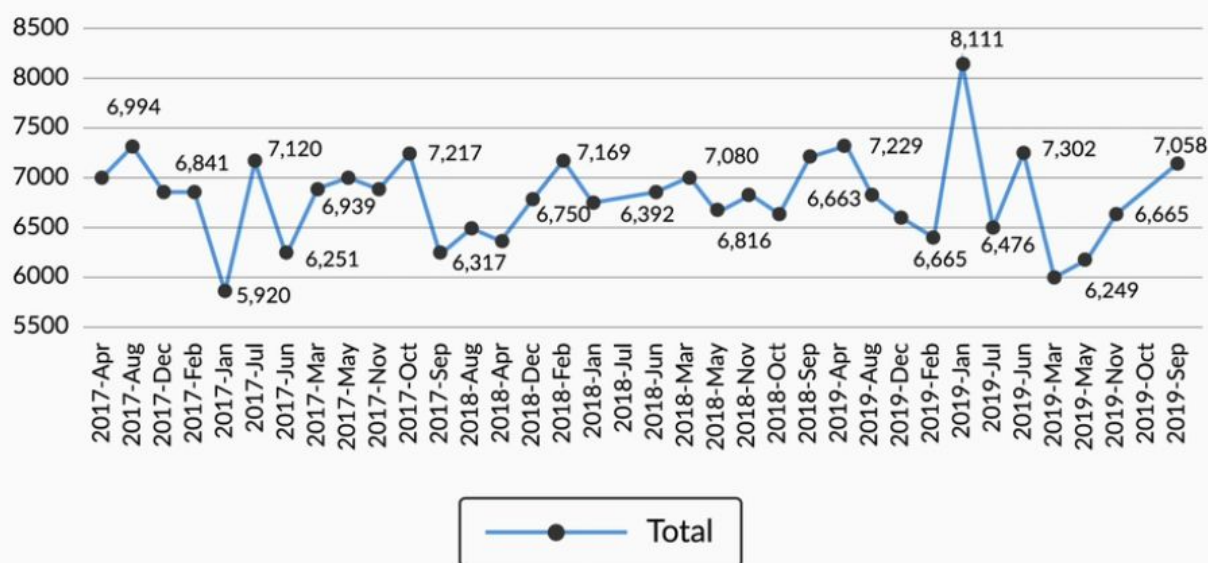
Instead of this, let’s try another approach to present the same information:

If we are presenting to the Head of Channel or to the Head of Distribution, then the title of the first slide would be ‘Our Top-Selling Product is Mostly Purchased Online’. This would allow us to talk to the Head of Channel or the Head of Distribution with regard to their respective business unit.



The second slide would also be quite insightful for the Head of Channel or the Head of Distribution.

INCREASE IN Q3 ONLINE ADVERTISING HAS IMPROVED SALES



As you can see, communicating insights in a manner that makes a connection with the audience ensures that we have their attention and they know that we are talking to them by being specific and relevant. And for that, we must keep our objective in mind.

So, in the the home loan example, assuming your objective is to provide information to the marketing head, the narrative would include the following:

- **The surface problem:** An effective graphic to tell them what is happening
- **Underlying causes:** Steady decline in the age of customers, probably due to a positive job market
- **Insights related to customer segments that are most likely to convert:** In other words, talking about what we have found after analysing the data
- **Implications of the trends**

However, if the objective is to influence decisions, then your narrative would include the following:

- **Key insights:** Underlying causes
- **Action needed:** The list may or may not be exhaustive

Note that the audience here may not bother much about the data, variables, processes, methodology or even the assumptions. And this is why it would be difficult for us, as data analysts, to not talk about the aspects on which we invested time.

Nevertheless, we need to remember that just like it is impossible to have a good movie without a good story, irrespective of how well it is told, it is impossible to rely on a good narrative if our insights are not strong. Without good insights, the quality of our narrative does not matter. It is, therefore, extremely important that you have informative, relevant and actionable content before you think about the narrative.

There is a famous saying for making great presentations: “***Tell your audience what you are going to tell them, then tell them, then tell them what you told them***”.

This means repeating the core message/objective of the presentation thrice. This helps us ensure that the content of our presentation is much more structured and also the flow of the topics is much more logical.

The saying above is trying to convey the following:

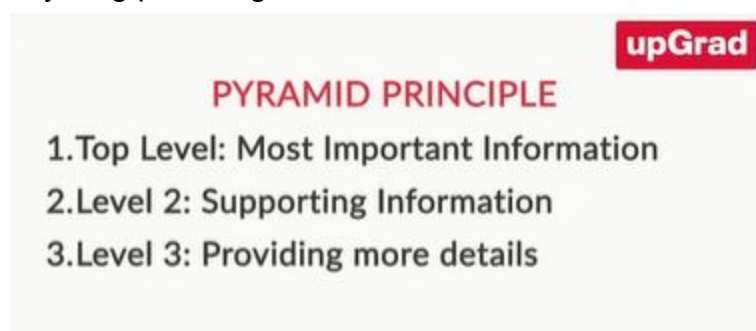
1. Explicitly state your agenda or the objective of your presentation to eliminate the element of surprise for the audience.
2. While covering various topics, you need to be crisp, clear and to the point.
3. Repeat the core message of your presentation, and finish with strong recommendations and conclusions

The most important part of a presentation is a good structure. To build a good structure, you should include the following elements in your presentation:

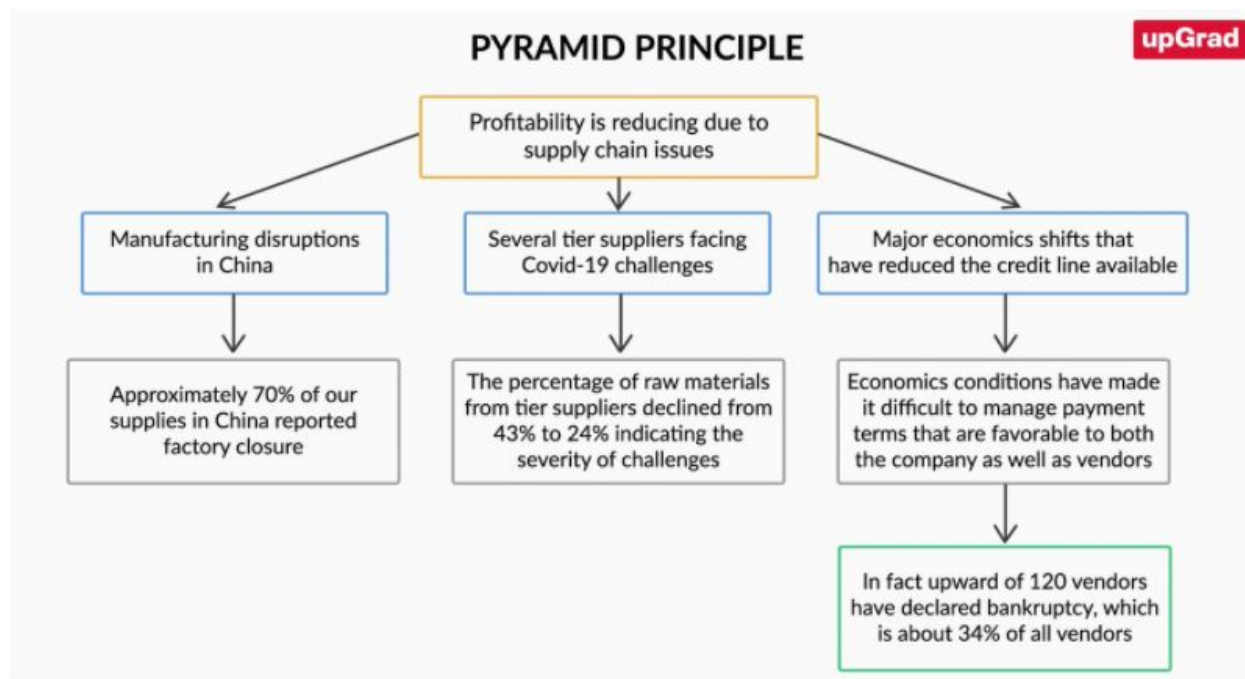
1. **Storyboarding:** It helps you create a visual outline of your presentation. Sequence does not matter here. What matters is that you think out loud by putting down all of your thoughts on the storyboard. While storyboarding, jot down points such as agenda, objective, key findings. It is advisable to not use any software while creating your storyboard. Instead, use a pen and paper, or a blank space that gives you the freedom to think while creating your storyboard.
2. **Foundational elements:** Always start with the foundational elements, which include Agenda, Objective, Main Body and Conclusion. They help you understand how much time it would take for you to deliver the presentation.



3. **Pyramid principle:** It allows you to keep your audience in sync with your content flow. It asks you to put the most important point first and explain the core objective and the insight of the presentation to the audience. While preparing your presentation, you must remember the most important information and then summarise everything providing more information.



Let's take an example to understand this better. Suppose a company XYZ Manufacturing is facing severe profitability issues. Your team has been tasked to analyse the underlying data, understand it and then explain what is going on, citing suitable reasons? You are also required to share your findings in the form of a presentation. The pyramid principle for this example would look like this:



4. **Main body:** This is the part that receives the most attention from the audience and, therefore, the content included here should be crisp and interesting. This part of a presentation is quite informative in nature, and elements such as frameworks, process charts and Venn diagrams are incorporated to display necessary information in an interesting yet understandable manner.

Nevertheless, always remember that you are not supposed to add details on the source of the insights that you are including in the main body of your presentation. References are included in the appendix section.

5. **Conclusion:** This is the final part of the structure of your presentation. No new information is presented in this section. You are expected to provide a summary of your presentation, highlighting the main points, and provide strong recommendations.

VISUALISATION

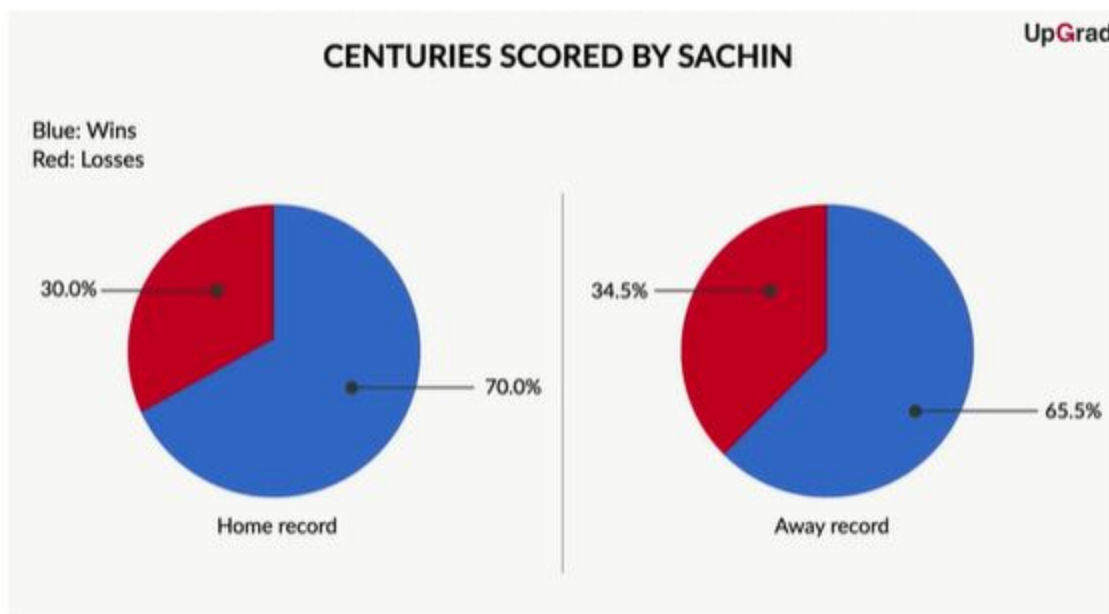
The term 'visuals' is different from visualisation, since the latter usually refers to the use of charts and graphics to analyse or explain data-based relationships. There is a large part of information that is non-quantitative in nature, or not related to data, but is related to critical components of your story. Visuals are used to represent such non-quantitative information in a more interesting manner.

Data visualisation helps make data much more understandable for the audience. According to some researchers, when data is presented using some visualisation, say graphs, a lot more people can understand and infer from it quickly, in contrast to the same information being presented in a tabular or a message format.

An example of this is shown below.

UpGrad

ODI CENTURIES SCORED BY SACHIN						
No.	Score	Pos.	Inn.	S/R	Venue	Result
39	100	2	1	88.49	Arbab Niaz Stadium, Peshawar	Lost[101]
40	141*	2	1	95.27	Kinrara Academy Oval, Kuala Lumpur	Lost[102]
41	100*	4	1	131.57	IPCL Sports Complex Ground, Vadodara	Won[103]
42	117*	1	2	97.5	Sydney Cricket Ground, Sydney	Won[104]
43	163*	2	1	122.55	AMI Stadium, Christchurch	Won[105]
44	138	1	1	103.75	R. Premadasa Stadium, Colombo	Won[106]
45	175	2	2	124.11	Rajiv Gandhi International Stadium, Hyderabad	Lost[107]
46	200*	2	1	136.05	Roop Singh Stadium, Gwalior	Won[108]
47	120	2	1	104.34	M. Chinnaswamy Stadium, Bangalore	Tied[109]
48	111	2	1	109.9	VCA Stadium, Nagpur	Lost[110]
49	114	2	1	77.55	Sher-e-Bangla National Stadium, Mirpur	Lost[111]



To gain insights from the data presented in the table given above, we have to go through the numbers, although with the pie charts, it takes us very less time to understand what the data means and what we should infer from it.

TYPES OF VARIABLES

1. Qualitative Variables

- **Nominal:** Used for labelling variables without any scale. There is no overlap or order. For example : Gender of a person.
- **Ordinal:** They are basically nominal variables with an underlying order. For example: Food ratings.

2. Quantitative Variables

- **Intervals:** Numeric measures where the underlying order and the difference is known. Ex: Temperature.
- **Ratios:** Numeric measures which are basically interval scale variables along with an absolute zero allowing several forms of statistical analyses to be performed. Ex: Height

VISUALISATION OF QUANTITATIVE VARIABLES

1. Scatter Plots

- Compares the relationship between two measures
- Useful for finding correlations
- Can be used to study data distribution within a specific interval using dot density

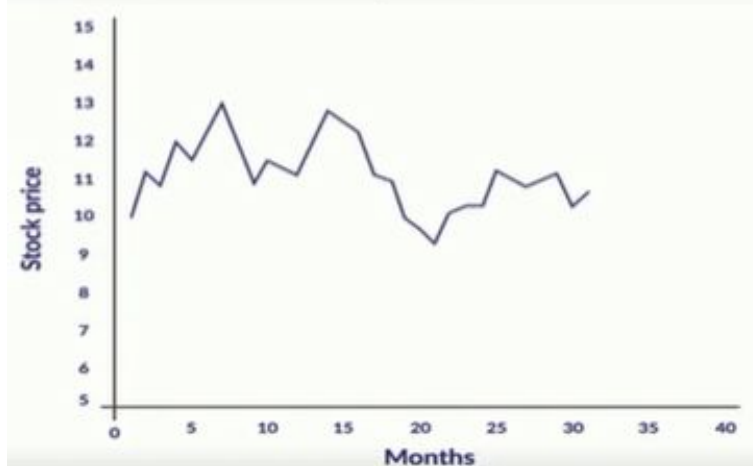
For example, observe the following scatterplot that shows the distribution of the salaries with respect to the ages of employees of a particular company.



2. Line Charts

- Represents the measure spread over a period of time
- Useful for time-series analysis, for example, for finding trends and showing relationships between variables

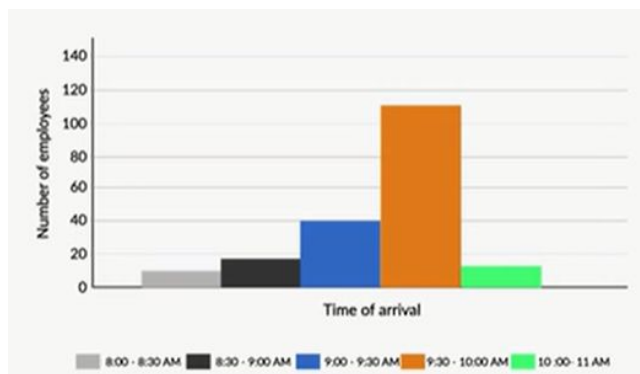
- The Y-axis shows a quantitative variable whereas the X-axis denotes a time period.



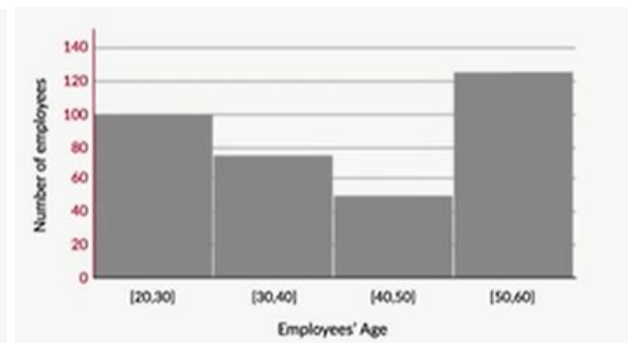
3. Histograms and Frequency Polygons

- Useful for visualising distribution of numerical data
- x-axis represents the variable of interest, which is divided into intervals and y-axis represents the frequency of occurrence in that interval

The image given below shows an example of a histogram.



ARRIVAL TIMES OF
EMPLOYEES

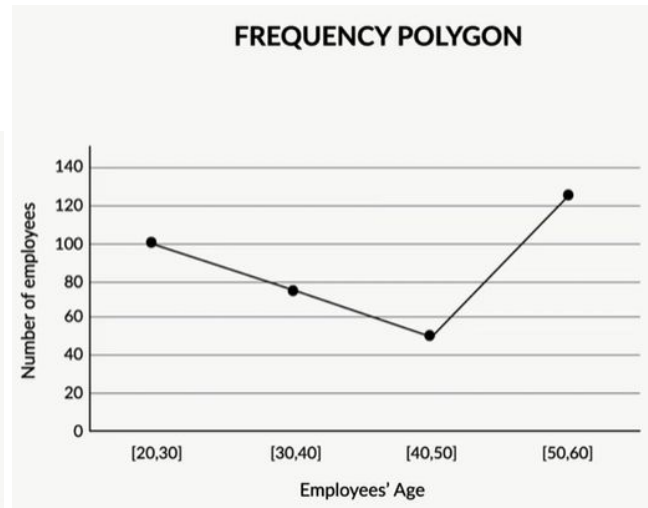
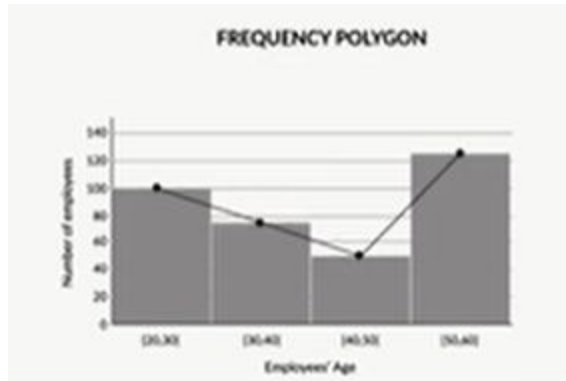


MOST OF THE COMPANY'S EMPLOYEES ARE IN
THE RETIREMENT AGE

4. Frequency Polygons

- Formed by joining the tips of a histogram
- Y coordinate represents the frequency in a particular interval

- Useful when two or more plots are compared, or when the histogram has a large amount of intervals, leading to clutter



VISUALISATION OF QUALITATIVE VARIABLES

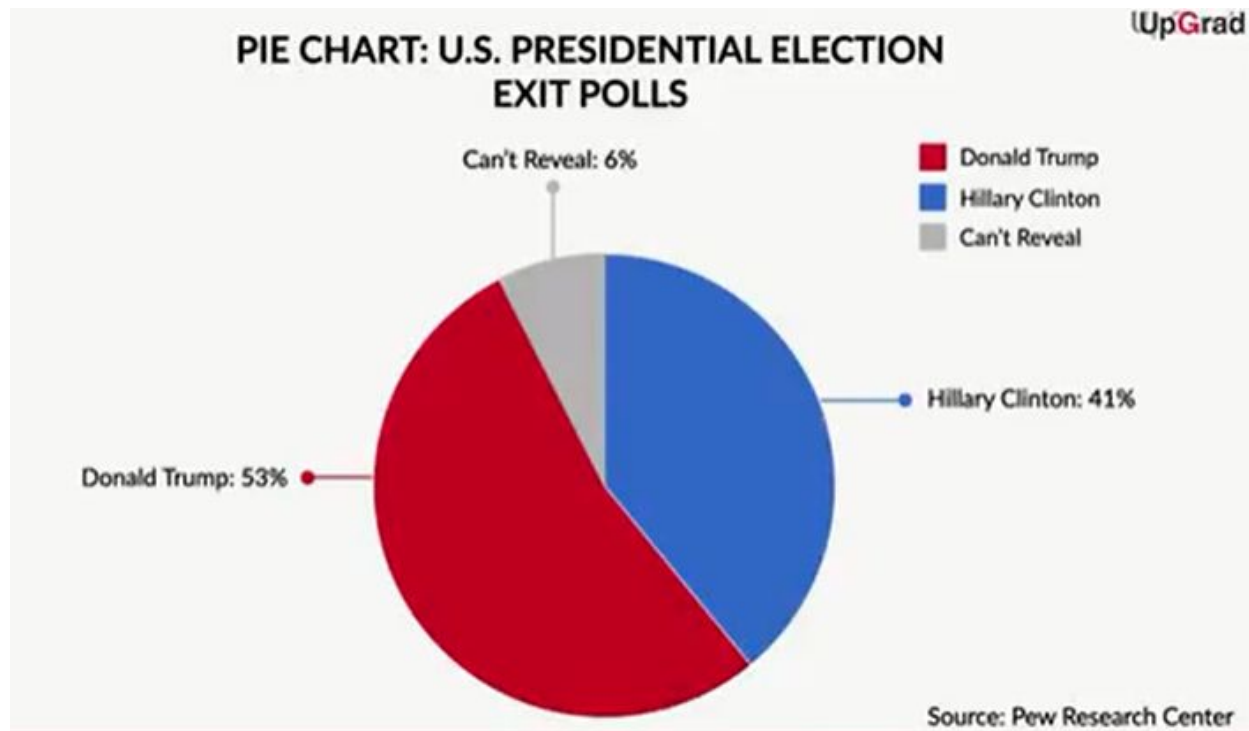
1. Pie Charts

- The most common method of representing categorical data. Each category is represented with a different colour.
- Useful for depicting part-to-whole relationships
- Not advisable to use for more than 6–7 categories as the viewers might not be able to identify them immediately

An example of a pie chart is shown in the image given below.

Let's see this example: According to one of the exit polls of US presidential elections 2016, 53% of the male population voted for Donald Trump and 41% of them voted for Hillary.

The remaining 6% didn't reveal who they voted for. So, the total number of male votes here are categorised into 3 categories – Donald Trump, Hillary Clinton, and those who didn't reveal.

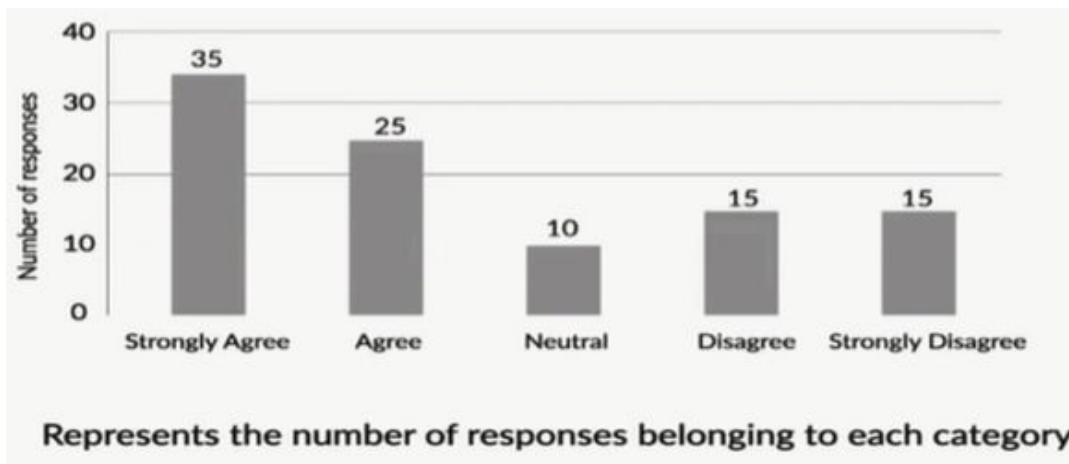


2. Bar Charts

- Mostly used for representing categorical data
- Useful when the number of categories is too high for a pie chart
- Rectangles in a bar chart represent the number of observations belonging to each category

An example of a bar chart is shown in the image given below.

In the following hypothetical survey conducted by Youtube, a bar chart shows the distribution for the types of responses received to the question "Do you think Youtube cares about users' privacy?"



VISUALISATION OF OTHER TYPES OF INFORMATION

1. **Word Clouds:** Used extensively when you need to show the frequency of a particular word in your analysis. These are usually popular in survey analysis, but are used in other cases as well.

For example, if you conducted a survey and want to show the distribution of the countries to which your respondents belonged, then this is how your word cloud could look, with the larger font for the larger sample sizes and the smaller font for the smaller samples.



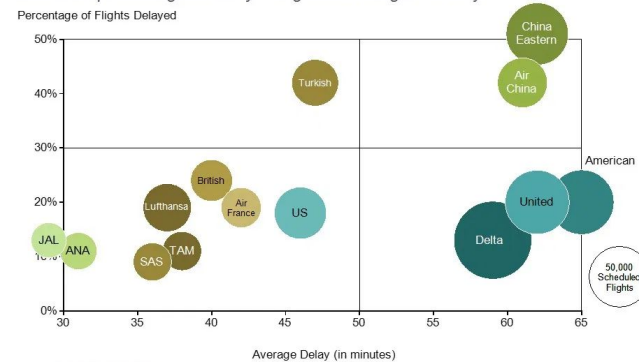
2. **Bubble Charts:** They are generally used to visualise quantitative information, but can also be used to visualise a combination of quantitative and qualitative information. When creating a bubble chart, you can show one variable on the x-axis and one on the y-axis, while the size of the bubbles is the third variable. To show a fourth dimension, you may use the colour or the pattern of the bubble, although it is generally used to show three variables.

Given below is an example of a bubble chart, which shows delays in airlines across the globe for 2015. On the x-axis, you can see the average delay in minutes, while on the y-axis, you can see the percentage of flights that were

delayed. The size of a bubble indicates how big an airline is. The number of scheduled flights was used as an indicator of size for this bubble chart. The colour of bubbles is not quite a fourth dimension; it is used to group the airlines by region.

Airline Delays Among Largest Global Carriers

Asian airlines have both the best and worst performance among the largest carriers in percentage of delayed flights and length of delay.



3. **Heat Maps:** They help you compare a large number of datapoints together. A good example of this is the use of a heat map by Phillip Cohen, a Professor of Sociology at the University of Maryland, to classify husband–wife couples according to their educational degree (majors) across 28 disciplines such as engineering, computer science, architecture and psychology.

The x-axis shows the man's major and the y-axis shows the woman's major. Green indicates a higher number of couples than expected for a particular combination of majors. Yellow indicates about as many couples as expected, while red indicates fewer couples than expected a particular combination of majors.

College major of wife and husband: College graduates married in the previous year

Couples in which women married for the first time only; 2009-2016 American Community Survey (N=27,806)

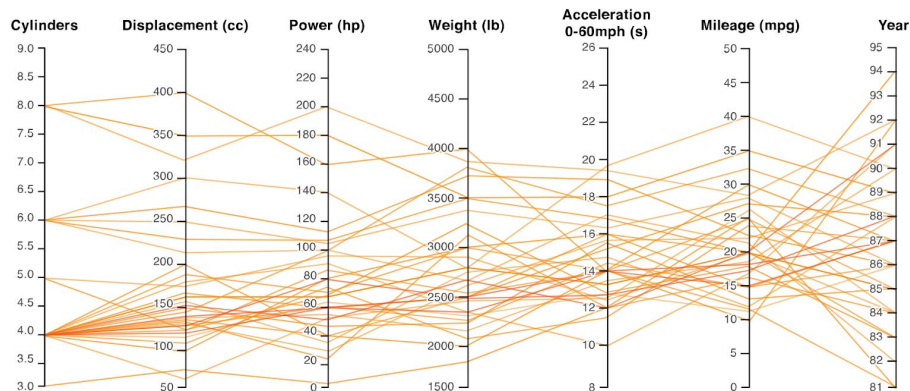
Philip N. Cohen

Ratio of observed to expected frequency

WIFE	HUSBAND																											
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28
1 Agriculture	19.6	3.4	.5	.0	.4	1.0	.6	1.1	.8	.0	5.3	.4	.0	1.2	.9	.0	.8	.0	.4	.7	.9	1.0	1.5	.4	.7	1.7	.7	.5
2 Environment NatRes	1.9	11.4	2.3	2.0	.9	.6	.4	.9	.2	.0	.0	2.4	.9	1.3	1.2	.0	.9	1.7	.0	1.2	.7	.5	.6	1.3	.6	2.3	.5	.6
3 Architecture	.2	.0	17.9	1.2	.4	1.2	.3	1.6	.0	.8	.0	.5	1.6	.8	1.5	.0	.5	2.7	.0	.3	1.1	1.1	.6	.8	1.0	.0	.7	.6
4 Area/Ethnic/CivStud	.7	1.7	2.2	6.9	1.3	1.5	.9	.6	.0	2.6	.0	1.0	.2	1.3	1.4	.6	.1	1.4	.0	1.0	.3	.0	.6	1.2	1.8	.6	.8	2.1
5 Communications	.7	.9	.4	1.2	2.6	.6	.8	.7	.8	.5	1.3	1.2	.9	.6	.9	1.1	.9	1.2	1.2	.7	.9	1.0	1.2	1.2	.8	.8	1.2	1.2
6 Computer/InfoSci	.4	.2	.4	.0	.8	4.7	.1	2.1	4.6	.1	.0	.3	.2	.5	.7	.4	.1	.5	.1	1.0	.4	.3	.7	.3	.4	.7	.4	.2
7 Education Admin	1.7	.9	.7	.7	.9	.8	3.3	.7	.9	.7	1.4	.8	1.3	.7	.8	1.0	1.2	.7	2.1	.6	.9	1.4	.4	.7	.9	1.0	1.0	1.3
8 Engineering	.2	.6	.9	.2	.4	1.8	.3	3.5	1.1	.5	.0	.3	.1	.6	.5	.4	.5	.2	.3	1.2	.2	.4	.9	.4	.4	.6	.4	.2
9 Engineering Techn	1.3	.0	.0	.0	.1	2.2	1.3	2.5	7.0	1.5	.0	.0	.0	1.1	.5	1.1	.3	.7	.7	.0	.2	.0	.0	.4	.8	1.4	.5	.0
10 Linguistics/Foreign	.2	1.6	1.2	1.2	.9	.9	.8	.9	.9	5.0	.0	1.9	.1	1.1	2.0	1.6	1.4	1.5	.7	.8	.7	.6	1.3	1.4	1.3	.4	.7	1.4
11 Family/Consumer Sci	.5	1.6	1.7	2.2	1.5	1.0	.5	1.0	.7	.5	2.5	1.3	2.4	.6	.8	1.5	2.5	.5	.3	.7	1.3	1.3	1.6	1.1	.7	.2	1.0	.6
12 English Language/Lit	.9	1.2	.8	1.4	1.3	1.1	.7	.7	.5	1.3	.4	2.9	1.6	.9	1.7	1.5	.6	.9	.9	.9	1.3	.6	.8	1.3	1.3	.4	.8	1.8
13 LiberalArts/Human	.3	.4	3.6	3.1	1.3	1.0	.8	.8	.6	1.2	.0	1.2	9.4	.7	1.6	1.3	.5	.2	1.0	1.5	.8	1.5	.3	1.1	.8	.8	.7	1.2
14 Biology and LifeSci	.9	1.2	1.2	.9	.6	.9	.5	1.0	1.0	.9	.8	.6	3.0	.8	.8	1.0	.7	.8	1.8	1.1	.8	.8	1.0	.6	1.2	.7	.8	
15 Math/Stats	.0	.3	.8	.0	.5	1.6	1.3	1.3	2.1	1.7	.0	1.3	.8	.6	5.7	1.7	.5	1.2	1.8	2.0	.5	.3	.8	.7	.9	1.1	.7	.5
16 Interdisc/multidisc	1.5	.5	.3	1.8	1.0	1.1	.6	.9	2.5	1.3	.0	.9	.8	1.6	.7	2.9	1.2	1.6	1.7	1.0	1.8	.5	.0	1.1	.9	.3	.8	1.2
17 PhysFit, Park/Rec	1.3	.7	.5	.4	.6	.5	1.3	.7	2.0	.8	1.6	.4	.6	1.1	.2	1.5	5.7	.5	.4	.5	.9	.8	2.5	.9	.5	.9	1.3	1.2
18 Philosophy/ReligStud	.9	1.5	1.2	1.4	.9	.7	.4	.9	1.0	3.7	.0	1.8	.7	.7	2.8	4.2	.4	8.3	1.6	.7	1.8	1.1	.0	1.0	1.2	.2	.6	1.9
19 Theology/ReligVoc	.0	.0	2.6	1.7	1.1	.3	2.4	.5	.4	1.5	.0	.4	.0	1.3	2.7	2.7	.0	4.0	31.0	.0	1.0	1.4	2.7	.3	1.7	1.5	.5	.4
20 PhysicalSci	.4	.8	.4	2.0	.4	.9	.6	.9	.5	1.9	.5	.8	.5	1.2	2.0	1.0	.6	.9	.2	7.2	.6	.6	.0	.9	.5	.8	.6	1.0
21 Psychology	.9	1.3	.8	1.0	1.0	.8	.8	.8	.4	.8	2.4	.8	.7	.8	1.0	1.4	1.0	2.0	1.2	.7	1.9	1.4	1.5	1.2	1.3	.8	1.0	1.3
22 CrimJustice/Fire	.6	.5	.5	1.5	1.0	.9	1.2	.5	.9	1.9	2.7	.2	2.3	.6	1.0	1.0	1.1	1.1	.3	.7	.5	6.3	1.1	.8	.9	1.8	1.2	.2
23 PubAff/Policy/SocWk	1.5	.6	1.1	3.9	1.0	1.1	.9	.8	1.1	1.9	2.0	1.4	1.3	.8	.3	.2	.4	.7	1.1	1.3	1.4	1.0	6.9	.9	.3	.6	1.2	.4
24 SocialSci	.4	.9	1.2	1.6	.9	.9	.5	.8	.4	.7	.3	1.2	1.1	1.0	1.1	1.4	.5	1.9	.3	1.0	1.4	.7	1.7	1.8	1.1	.4	.9	1.3
25 Fine Arts	.5	1.0	1.6	1.1	1.4	1.0	1.0	.7	.6	1.2	.5	1.8	1.3	.4	.9	.9	.8	1.0	1.1	.7	.7	.8	.6	1.0	4.4	.5	.8	.8
26 Medical/HealthSci	.8	1.1	.5	.7	.8	.7	1.1	1.0	1.0	.8	1.7	.6	.9	1.4	.8	1.0	1.6	.9	.7	.8	1.2	1.2	1.0	.7	.5	3.4	1.0	.9
27 Business	1.0	.7	.7	.5	.9	1.0	.6	1.0	1.3	1.1	.8	.7	.6	.7	.8	.5	.8	.3	.7	.6	.8	1.1	.7	.9	.6	.7	1.6	.6
28 History	.4	1.1	1.5	.5	.8	.7	1.0	.7	.2	.6	.7	2.0	.4	1.0	1.0	.8	.8	2.9	1.0	.8	1.2	.9	.7	1.6	1.0	.6	.8	2.9

4. **Parallel Coordinates Plots:** This is yet another visualisation technique that is used when we need to show multivariate, numerical data. In a parallel coordinates plot, each variable is allotted its own axis and all the axes are parallel to each other. The variables are generally arranged according to the way how the user understands the data.

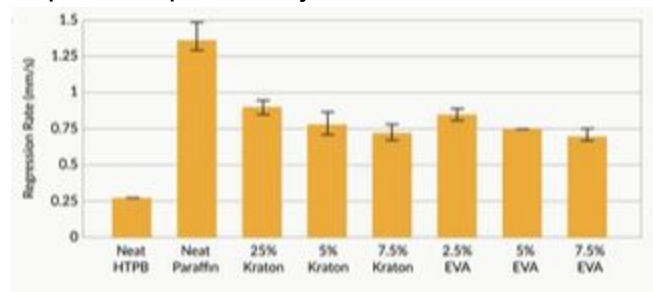
For example, if you had to plot multiple variables related to cars, such as cylinders, displacement, power, weight, acceleration, mileage and year of manufacture, then you may use a parallel coordinates plot.



VISUALISATION OF ADVANCED ANALYTICAL TECHNIQUES

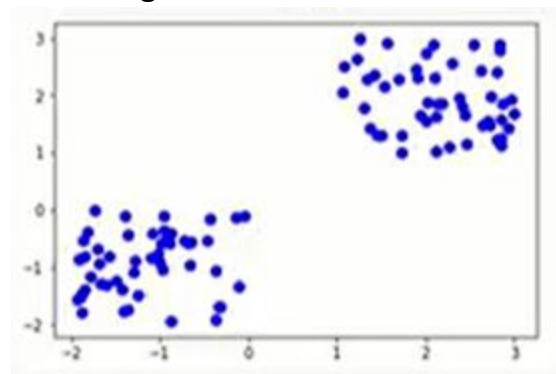
Regression Analysis: After performing a regression analysis, you may not want to present your data using multiple line graphs as they might not convey the data effectively. For presenting the data properly, you might have to show the final set of variables that you found were the most impactful. This is where impact analysis comes into picture.

Here is an example of impact analysis>



Similarly, to show the results of a clustering algorithm, you don't need to show all of the clusters; instead, you need to show only the 3–4 clusters that you have driven.

K-means or Clustering



VISUAL DESIGN PRINCIPLES AND STORYBOARDING

Now, you will learn how to design your visuals intelligently so that they capture maximum attention.

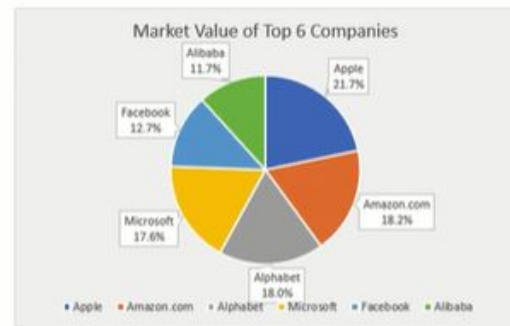
Accuracy It is the quality or state of being correct.

- Correctness of visuals
- How accurately are you describing your charts

Precision It is the quality or state of being exact.

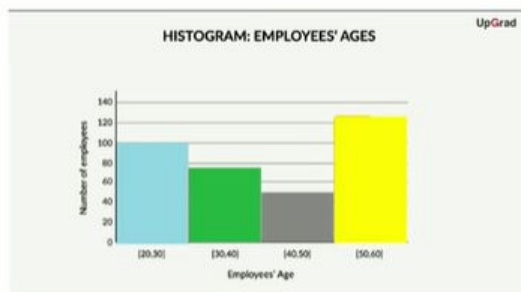
- Level of details in the visuals to present data better

PRECISION

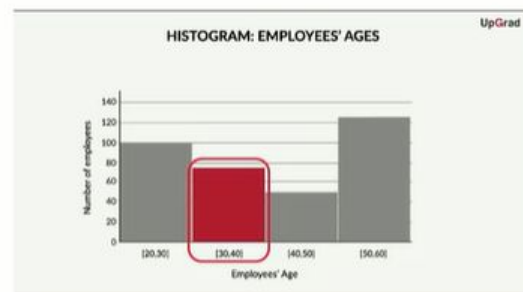


RESTRICT THE USE OF DECIMALS UPTO 2 POINTS

DRAWING ATTENTION



LIMIT THE USE OF VISUAL DIFFERENCES TO HIGHLIGHT ONLY SIGNIFICANT DIFFERENCES



GUIDE AUDIENCE'S ATTENTION TO MORE DISTINCT & SALIENT FEATURES

Drawing attention with Text: Use preattentive features to highlight important sections.

1. Applying bold formatting to text. 2. Using different fonts. 3. Italicising text. 4. Using different colours. An example of this is shown in the image given below.

DRAWING ATTENTION

With Text

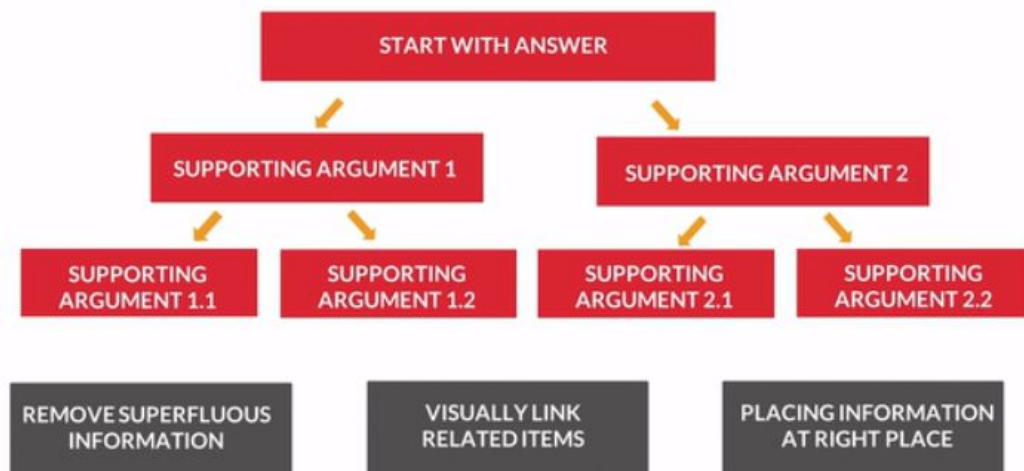
- India won **70%** of times when Sachin Tendulkar hit a century.
- Internet of Things (IoT) will be the trend, which will generate more than **\$300 billion** annually by 2020.
- Almost 25-50% of data science teams are employing *Agile methodologies*, states Wikibon community research.
- Harvard Business Review states experience economy is where *"a company intentionally uses services as the stage, and goods as props, to engage individual customers in a way that creates a memorable event."*

STORYBOARDING

1. Helps to assimilate information
2. Clarifies the logic of your hypothesis through supporting arguments
3. Helps identify gaps in the analysis
4. Helps avoid redundant work

An example of storyboarding is shown in the image given below.

STORYBOARDING



GOLDEN RULES OF DATA STORYTELLING

Here are the five golden rules that can help you achieve effective storytelling in your presentations:

1. **Present insights, not data:** The first rule of data storytelling is to make insights and not data the focus of each slide, and in fact the entire presentation.
2. **Know what not to present:** Your aim must be to present only what is absolutely critical for the audience to know at that time.
3. **Use variation (in colour, shape, size, style) only** to highlight importance and relevance, and not to show appeal or aesthetics

4. **Speak the business language: 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.

For example, say you conducted a correlation analysis between loan amount applied for and gender. An excellent method of communicating the result would be to simply state it in layman terms than to say that the coefficient of correlation between these two variables was high.

5. **Declutter**

It is important to ensure that your presentation is free from clutter. To do this, you can use the whitespace attribute, which is often overlooked in presentations. As data analysts, we often look at our slides in terms of the amount of information that we include, or our communication, and focus less on the amount of space that we leave. White space allows viewers to focus on the most important points only and draws their attention away from noise.

6. **Rule of three**

A very effective piece of communication advice that works well not just while presenting but also while writing in general is the rule of three. It states that “if you want something stuck in someone’s head, put it in a sequence of three”.

Disclaimer: All content and material on the UpGrad website is copyrighted material, either belonging to UpGrad or its bonafide contributors and is purely for the dissemination of education. You are permitted to access print and download extracts from this site purely for your own education only and on the following basis:

- You can download this document from the website for self-use only.
- Any copies of this document, in part or full, saved to disc or to any other storage medium may only be used for subsequent, self-viewing purposes or to print an individual extract or copy for non-commercial personal use only.
- Any further dissemination, distribution, reproduction, copying of the content of the document herein or the uploading thereof on other websites or use of content for any other commercial/unauthorized purposes in any way which could infringe the intellectual property rights of UpGrad or its contributors, is strictly prohibited.
- No graphics, images or photographs from any accompanying text in this document will be used separately for unauthorised purposes.
- No material in this document will be modified, adapted or altered in any way.
- No part of this document or UpGrad content may be reproduced or stored in any other web site or included in any public or private electronic retrieval system or service without UpGrad's prior written permission.



- Any rights not expressly granted in these terms are reserved.