



BITS Pilani
Pilani | Dubai | Goa | Hyderabad

Introduction to Data Science

Storytelling with Data – The Final Deliverable

Dr. Ramakrishna Dantu

Associate Professor, BITS Pilani

Introduction to Data Science



Disclaimer and Acknowledgement



Disclaimer

- The content for these slides has been obtained from books and various other source on the Internet
- I here by acknowledge all the contributors for their material and inputs.
- I have provided source information wherever necessary
- I have added and modified the content to suit the requirements of the course

Introduction to Data Science



Storytelling with Data – The Final Deliverable

- The Deliverable
- The Narrative
- The Report Structure
- What is a final deliverable?
 - It is not merely a deck of PowerPoint slides with tables and graphics
 - It is the narrative or the story that the data scientist tells from the insights presented as graphs and tables



The Final Deliverable



What makes one a Data Scientist

- What sources say about data scientist/data science?
 - a data scientist is someone who specializes in data collection, storage, and manipulation
 - a data scientist as the one who focuses on computer engineering
 - and the engineering aspect of data management overrides the science aspects of it and vice versa
 - data science is all about software and coding
 - some data scientists focus on the size of data and others would focus on what to do with the data
 - data science is all about forecasting and predictive analytics
 - data science is more to do with data mining
- However, seldom will you find anyone stress the importance of the narrative

The Final Deliverable



What makes one a Data Scientist?

- A data scientist should be as concerned with the **narrative** as with data and analytics
- "A data scientist is a **storyteller** who weaves **narrative** from bits and bytes, thus presenting a **comprehensive argument** to support the evidence-based strategic planning, which is essential in a data-rich world awash with analytics."
 - Murtaza Haider
- despite the "compelling reasons for the importance of **stories**, most quantitative analysts are not very good at creating or telling them. The implications of this are profound—it means analytical initiatives **don't have** the impact on **decisions and actions** they should. It means time and money spent on acquiring and managing data and analyzing it are effectively wasted."

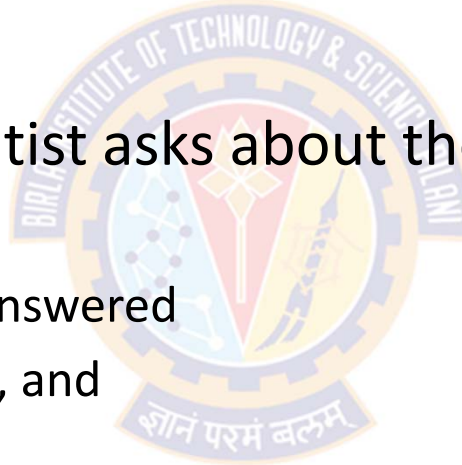
-- Thomas Davenport

The Final Deliverable



What makes one a Data Scientist?

- A good data scientist builds a **comprehensive road map** before he/she embarks on the journey
- The questions a data scientist asks about the final deliverable before embarking on analytics
 - What questions need to be answered
 - What resources are available, and
 - What stories need to be told



The Final Deliverable



Ingredients

- The ultimate purpose of analytics is to **communicate findings** to the intended audiences
- Analytics summarize findings in **tables and plots**
- These summary tables, figures, plots, etc., must be complemented with a **narrative**
- The stakeholders might use these insights or findings to formulate **policy or strategy**
- In academia, the final deliverable is in the form of **essays and reports**.
 - Such deliverables are usually 1,000 to 7,000 words in length.

The Final Deliverable



Report Example

- Good example for a deliverable that builds narrative from data and analytics
 - [United States Economic Forecast](#)
- The 24-page report focuses on the state of the U.S. economy as observed in December 2014



The Final Deliverable



The Report

- opens with a "grabber" highlighting the fact that contrary to popular perception, the economic and job growth has been quite robust in the United States
- is not merely a statement of facts, it is a carefully crafted report that follows a distinct theme
- focuses on the "good news" about the U.S. economy, which include:
 - the increased investment in manufacturing equipment in the U.S. and
 - the likelihood of higher consumer consumption resulting from lower oil prices.
- uses time series plots to illustrate trends in markets
- has the accompanying narrative that focuses on income inequality in the U.S. and refers to Ira Kalish's article on the same
- has other graphics focused on housing, business and government sectors, international trade, labor and financial markets, and prices
- The appendix carries four tables documenting data for the four scenarios discussed in the report

The Final Deliverable



Example

- Before the authors started their analysis, they
 - must have discussed the **scope** of the final deliverable
 - would have deliberated the **key message** of the report
 - then would have looked for the **data and analytics** they needed to make their case
- The initial **planning** and **conceptualizing** of the final deliverable is extremely important for producing a **compelling document**
- Without due consideration to the final deliverable, is likely to result in a poor quality document where the **analytics and narrative** would **struggle to blend**

The Final Deliverable



Example

- Before embarking on analytics, the following questions must be considered:
 - What is the research question
 - What answers are needed (Findings)
 - How have others researched the same question in the past? (Literature Review)
 - What information do you need to answer the question? (Data Collection)
 - What analytical techniques/methods do you need? (Methodology)
 - The Narrative (Results and Discussion)
 - The Report Structure

The Final Deliverable



The Research Question

- The key to good analytics is having the research question
 - that is new, interesting, relevant
- All analytics are conducted to gain insights into a research question
- A better understanding of the research question ensures that the subsequent actions are likely to bear fruit
- It is also important to recognize that advanced analytics are **not a panacea** for a poorly structured research question

The Final Deliverable



The Research Question – Example

- You are a senior director for human resources
- You are asked to review a grievance filed by women workers who contend that the firm pays them less than what it pays men
 - Your goal is to investigate **whether a gender bias exists in compensation**
- If you are asked to run a query on the average compensation for each department broken down by gender, then the appropriate question would be:
 - **Does compensation differ for men and women employees in the firm?**
- Gender bias in this particular case implies that women employees are paid less for doing the same job when no difference exists in **experience**, **education**, **productivity**, and **other attributes** of the work done by men and women.
 - The research question thus **has to account for relevant attributes**
- A more appropriate research question is stated as follows:
 - **Given the same education, experience, and productivity levels, does the firm compensate women less than men?**

The Final Deliverable



What Answers Are Needed?

- After refining the research question, need to think about the answers or solutions that we will have to present to the stakeholders
 - Think not merely about the **format** of the final report, but also about the **content** in the answers
- If we reach a rather fuzzy conclusion that suggested the gender bias **might have** taken place in some circumstances but not in others
 - This is not going to help executives to take any action
- Building on the gender bias example, the executives might want your report to offer **robust evidence** that either **refutes or confirms** the allegations
 - must choose those **methods and data** that will help to reach a conclusive answer
 - must also be cognizant of the **deadlines**
- **The analytics should inform and guide the narrative and not the other way around**
- Think of the format in which the answer must be presented to the executives
 - If they are expecting a **report or a PowerPoint presentation**, then produce one that will help to convey the findings

The Final Deliverable



How Have Others Researched the Same Question in the Past?

- After sharpening our research question, we might want to see how others in the past have researched the same or similar question
- In academia
 - This is referred to as a literature review
- In business
 - This is called background research
- This enables us to benefit from the experience of other researchers
- Need to consider the Sources of Information
 - Google Scholar, Web of Science, ProQuest Digital Library, IEEE, etc.



The Final Deliverable



What Information Do You Need to Answer the Question?

- A review of relevant research will put us in a better position to determine what information we need to answer the research questions
- In the gender bias in wages example,
 - we would need the details such as compensation received by each employee, employees' sex, age, experience, education, job classification, and some measure of productivity, etc.
- If the firm did not maintain detailed records, it will limit the analyst's ability to answer the question about gender bias
- **No amount of technical savviness can be a substitute for little or poor information**
- Need to determine whether we can collect the information **internally or ask an external agency** to collect it
- Also need to consider whether the data is **latest or not**
- Need to take a call if the ideal information needed for the analysis might take too long or cost too much to obtain

The Final Deliverable



What Analytical Techniques/Methods Do You Need?

- It is important to know in advance what methods or techniques to use for the analysis
 - This has an impact on the type of data you need for the analysis
- Consider our gender bias example:
 - Need to determine whether a gender bias in wages exists when men and women with similar education, experience, and productivity are employed to do the same job
 - Analysts need to hold other factors constant while investigating the impact of gender
 - It appears that a regression-type model is better suited to answer the research question
- Also need to explore if other suitable alternative are available or not
 - If yes, what are the costs associated with employing them

The Final Deliverable



The Narrative

- The narrative gives life to numbers
 - Suspended in columns and rows, numbers in tables **seldom speak** to the reader
 - The **narrative** is always what calls attention to the **gems latent in tables**
 - Numbers slowly start to make sense as the speaker explains what they represent and imply
 - Some exceptionally talented researchers are capable of generating self-explanatory tables
- Even when the tables and charts are well-illustrated the need for powerful narrative never diminishes!
- Narrative
 - adds color to facts
 - allows you to contextualize the findings

The Final Deliverable



The Report

- Think about the structure of the report
 - Will it be a brief **report** or a longer **document**
 - A brief report drafted as **commentaries** on current trends and developments that attracted public or media attention
 - A detailed report offered a **critical review** of the subject matter with **extensive data analysis** and commentary
- The structure depends on the length of the document
 - A brief report is more to the point and presents a summary of key findings
 - A detailed report incrementally builds the argument and contains details about other relevant works, research methodology, data sources, and intermediate findings along with the main results.

The Final Deliverable



The Report Structure

- The deliverable should follow a format including
 - The cover page
 - Table of contents
 - Abstract or Executive summary
 - Introduction
 - Literature Review
 - Methodology
 - Results
 - Discussion
 - Conclusion
 - References
 - Acknowledgements
 - Appendices



The Final Deliverable



Have You Done Your Job as a Writer?

- A data scientist is expected to do a thorough analysis with the appropriate data, deploying the appropriate tools
- A **writer** is responsible for communicating the findings to the readers
- A checklist
 - Have you told readers, at the outset, what they might gain by reading your paper?
 - Have you made the aim of your work clear?
 - Have you explained the significance of your contribution?
 - Have you set your work in the appropriate context by giving sufficient background to your work?
 - Have you addressed the question of practicality and usefulness?
 - Have you identified future developments that might result from your work?
 - Have you structured your paper in a clear and logical fashion?

The Final Deliverable



Reference

- Getting Started with Data Science – Making Sense of Data with Analytics
– Murtaza Haider





Thank You!