4/29/2024

Communicate our results:

1. Know our audience
   In a professional setting this might include a team manager, business stakeholders, data/engineering team, public
   Tailor your communication appropriately
   Technical jargon may confuse the public
   Over-explaining to your manager could suggest you think he's stupid

2. Target your presentation – be clear and complete
   a. Your understanding of the problem
   b. Measure the impact (if possible)
   c. What data was available to make the model/do analysis
   d. Initial solution or hypothesis
   e. Solution/model – examples and visualizations
   f. Impact and state a clear action, where appropriate put into context
   g. Discuss future avenues of research

3. Avoid common errors:
   a. Don't gloss over key assumptions, or omit them
   b. Don't recycle presentations intended for different audiences
   c. Don't show visualizations without context
   d. Don't save all the insight until the very end – processes are iterative and piecewise
   e. Include an abstract or executive summary with key findings at the top
   f. Include an appendix at the back with additional technical details and supporting materials (mathematical equations, literature review, code examples, etc.)
   g. Invite feedback

Tell a story with your data

Use a narrative framework
Personalize to your target to your target audience if possible
Give evidence that you are solving a problem that exists and can be fixed by your analysis
Don't expect charts to speak for themselves
Use empathy with the audience to guide you:
    If you were in the position of the audience, what would you want to know?
    What information would you find persuasive?
Bring your experience to bear when possible.
Case studies may help to illustrate difficult points
It may help to find new ways to present data that are more compelling than past methods – use more than one method

What is interesting as an audience member might not necessarily be interesting to an analyst
Figure out the end before figuring out the middle
Write it down, fix it later
What is the essence of the story? What is the most economical way of telling it? Build from there

Technical information should be appropriate to the audience
Don't describe your results as a sequence of activities – this information could be included in an appendix
Don't include unnecessary graphics just because they are cool – all graphics should serve a purpose

Interactive graphics can be especially for time series data (animations)
Games can be used to illustrate more complex interactions of variables

How could your outcomes be included in actions?

Dashboards
- Present data interactively, users can manipulate it themselves
- East to read and interpret

Time series forecasting/models
- Trends
- Seasonality (predictable, short-term cycles)
- Cyclical components (long-term cycles, ~decades)
- Noise (error)

Types of forecasts:
- Naïve – uses the most recent values to predict the next value
- Averages – mean of historical data
- Seasonal naïve – repeats the values of the previous seasonal cycle

Regression-based analysis –
Autoregression (predicts future data based on one ore more preceding values, lag), and traditional regression, ARIMA (auto regressive integrated moving average)

Smoothing methods – moving averages/exponential smoothing

For train/test split, don't do it randomly. Instead split it in time. Use the first 80% as the training data, and the final 20% as the test data.

Resources:

1. https://medium.com/comet-ml/a-data-scientists-guide-to-communicating-results-c79a5ef3e9f1
2. https://medium.com/comet-ml/a-data-scientists-guide-to-communicating-results-c79a5ef3e9f1
3. https://www.datapine.com/blog/data-dashboards-definition-examples-templates/
4. https://otexts.com/fpp2/data-methods.html
5. https://www.datarobot.com/platform/mlops/?redirect_source=algorithmia.com

Extended commentary:

When communicating the results of data analysis, there are several things to consider:

Audience: Consider who the intended audience is and how technical or non-technical they are. Use language and visuals appropriate for the audience.

Goals: Be clear on the goals of the analysis and ensure that the results are relevant and meaningful to the intended audience.

Visualizations: Use appropriate visualizations such as charts, graphs, and tables to present the results. Ensure that the visualizations are clear and easy to understand.

Context: Provide context for the results by explaining the data sources, the methods used for analysis, and any limitations or assumptions made during the analysis.

Uncertainty: Acknowledge and communicate any uncertainty or limitations in the analysis, such as missing data or assumptions made during the analysis.

Actionable insights: Clearly communicate any actionable insights or recommendations that can be drawn from the analysis.

Ethical considerations: Consider any ethical considerations related to the data and analysis and ensure that the communication of results is done in an ethical and responsible manner.

By considering these factors when communicating the results of data analysis, you can ensure that your audience understands the insights and can make informed decisions based on the results.

It is important to know your audience because it helps you tailor your message and presentation to their needs, interests, and understanding. If you understand your audience, you can anticipate their questions, address their concerns, and provide them with relevant information in a way that resonates with them. By doing so, you can increase the effectiveness of your communication and make sure that your message is received and understood. Additionally, understanding your audience can help you avoid using jargon or technical language that they may not understand, which can lead to confusion or misinterpretation of the information you are presenting.

Here are some common errors in communicating data analysis results:

Technical Jargon: Using technical language or jargon that your audience may not understand can make your message difficult to understand. It is important to communicate in plain language that your audience can understand.

Lack of Context: Failing to provide the necessary context to explain the significance of the findings can lead to misunderstandings. Make sure to provide the necessary background and context to help your audience understand the data and its implications.

Overcomplicating the Message: Overcomplicating the message can lead to confusion and misinterpretation of the results. It is important to simplify the message and focus on the key findings.

Cherry-picking Data: Selecting only the data that supports your argument or ignoring data that contradicts it can be misleading. It is important to present all of the data and the limitations of the analysis.

Lack of Visual Aids: Failing to use visual aids such as charts and graphs can make it difficult for your audience to understand the data. Visual aids can help to simplify complex information and make it easier to interpret.

Lack of Actionable Insights: Presenting data without providing actionable insights or recommendations can be frustrating for your audience. It is important to provide clear and actionable recommendations based on the data.

Poor Design: Presenting data in a confusing or unattractive manner can make it difficult to read and understand. It is important to use a clean and well-designed presentation format that is easy to read and understand.

Telling a story with data involves more than just presenting numbers and charts. It involves crafting a narrative that communicates a clear message, engages the audience, and inspires action. Here are some tips on how to tell a story with data:

Start with a clear message: Before diving into the data, it's important to have a clear idea of what you want to communicate. This will help guide your analysis and ensure that your story stays on track.

Know your audience: Understanding who you're speaking to will help you tailor your message and choose the most effective way to present your data. Consider factors like their level of technical knowledge, their interests, and their goals.

Use visuals: Visuals can help bring your data to life and make it easier to understand. Choose charts and graphs that are appropriate for your data and use color, size, and labels to highlight key points.

Tell a story: Weave your data into a narrative that engages the audience and makes them care about what you're presenting. Use anecdotes, examples, and real-world scenarios to illustrate your points and keep your audience engaged.

Be concise: Keep your story focused and to the point. Avoid jargon and technical language that might confuse your audience, and use simple language to communicate complex ideas.

End with a call to action: After presenting your data and telling your story, be sure to leave your audience with a clear call to action. What do you want them to do with the information you've presented? Make sure your message is actionable and provides a clear path forward.

Time series analysis involves analyzing data that is collected over time, with the goal of identifying patterns, trends, and relationships that may exist between different variables. Some common time series analysis methods include:

Moving averages: This method involves calculating the average of a set of values over a specific time period. Moving averages are used to smooth out fluctuations in the data and to identify trends.

Exponential smoothing: This is a time series forecasting method that assigns more weight to recent data points and less weight to older data points. The result is a forecast that responds more quickly to changes in the data.

Autoregressive Integrated Moving Average (ARIMA): ARIMA models are used to forecast future values of a time series based on its past values. This method takes into account the trend, seasonality, and noise in the data.

Seasonal decomposition: This method breaks down a time series into its trend, seasonality, and residual components. By examining these components separately, analysts can gain a better understanding of the underlying patterns in the data.

Fourier analysis: This method involves decomposing a time series into its underlying frequencies using the Fourier transform. This can be useful for identifying periodicities and other cyclical patterns in the data.

Granger causality: This method is used to determine whether one time series is causing changes in another time series. It can be useful for identifying causal relationships between variables.

Vector autoregression (VAR): This is a multivariate time series forecasting method that models the joint behavior of two or more variables over time. VAR models can be used to forecast the future values of each variable in the system.

These are just a few examples of the many time series analysis methods that are available. The choice of method will depend on the specific nature of the data and the questions being asked.