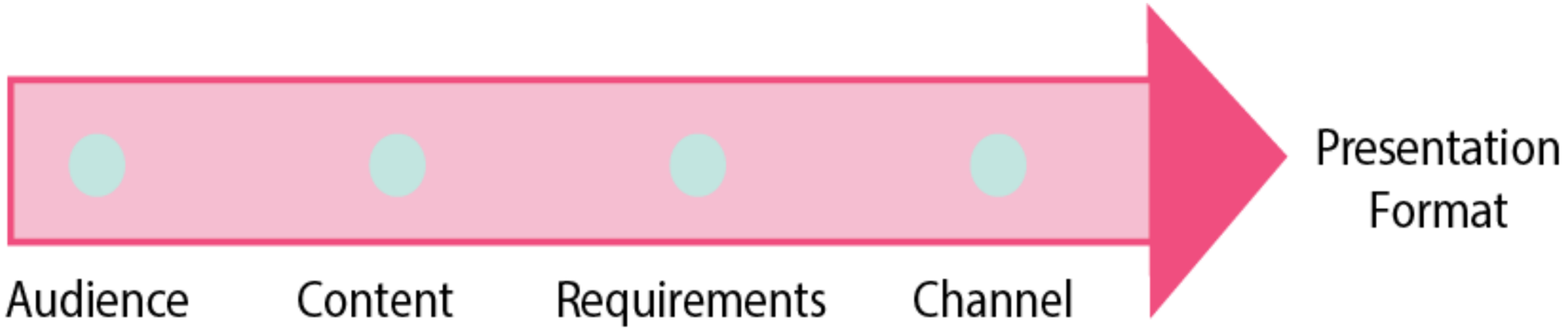
Types of reports

DATA COMMUNICATION CONCEPTS

Hadrien Lacroix
Curriculum Manager

Presentation strategy



Chapter 3

How to structure a written report?

- Types of reports
- Reproducibility
- Write precise and clear reports

Written reports

- Explain data analysis project
  - Sentiment analysis on product reviews

Written reports

- Explain data analysis project
  - Sentiment analysis on product reviews
- Communicate findings
  - 30% negative ratings for delayed shipping
  - Predict ratings with 90% accuracy
- Standards

Written reports

- Explain data analysis project
  - Sentiment analysis on product reviews
- Communicate findings
  - 30% negative ratings for delayed shipping
  - Predict ratings with 90% accuracy
- Standards
- Give recommendations to drive change

Types of reports

Informational

- Factual information
- Short
- Not strict structure
- Inform about facts

Analytical

- Analysis (relationships/recommendations)
- Varies (short or long)
- Strict structure
- Data-driven decisions

Final report

Elements

- Data analysis
- Findings and results
- Visuals

Format

- Long

Audience

- Details

Summary report

Elements

- Key findings and recommendations
- Visuals

Format

- Short (< 5 pages)
- Summary of final report
- Link to main document

Audience

- No need for details

Report structure

- Introduction
  - Purpose
    - Analysis of the product reviews gathered from website
    - Rating prediction based on review
  - Contextual information
    - Increase in negative reviews
  - Question of analysis
    - Factor affecting bad user experience

Report structure

- Introduction
- Body
  - Data
    - Description and tables
  - Methods
    - NLP and Random Forest
  - Analysis
    - Visuals
      - Graphs with most common words
  - Results
    - Description and visuals
      - 30% negative ratings associated with words "delayed" and "shipping".

Report structure

- Introduction
- Body
  - Data
  - Methods
  - Analysis
  - Results
- Conclusions
  - Restate question
  - Summarize important results
  - Add recommendations

Report structure

- Business context
- 1-3-25
  - 1 page of abstract
  - ? 3 pages of executive summary
  - ? 25 pages of detail

Audience

- People with little time
  - Introduction
  - Conclusion
  - Scan body

Audience

- Customer or internal collaborator
- Executive team
  - Scan introduction and conclusions
  - Recommendations

Audience

- Customer or internal collaborator
- Executive team
- Technical stakeholder
  - Body

Let's practice!

DATA COMMUNICATION CONCEPTS

Reproducibility and references

DATA COMMUNICATION CONCEPTS

Hadrien Lacroix
Curriculum Manager

Written report

A report must be clear and **reproducible.**

Reproducibility example

- Baking a cake
  - Recipe
  - Raw ingredients
  - Our oven and kitchen measuring gadgets
  - Cake with a **similar flavor**
- Data project
  - Run analysis again - **same results**

Replicability example

- Baking a cake
  - Own utensils
  - Own ingredients
- Data project
  - Different environment

Reproducibility and replicability virtues

- Prevents duplication of effort
- Build upon pre-existing work
- Focus on new challenges
- Peer review
- Tool agnostic

Best practices

1. Keep track of how results were produced
    - Well document scripts
        - Comments in code
    - List packages and environment used
    - Version control

Best practices

1. Keep track of how results were produced
2. Avoid manual data manipulation
    - Data versioning
    - Store raw data and intermediate steps
    - Adapt and resolve problems
    - Example: data imputation
        - impute missing values with the mean
        - save and close editor
        - how to know which values were replaced in the first place?

Best practices

1. Keep track of how results were produced
2. Avoid manual data manipulation
3. Control randomness
    - Random seeds for ML pipelines
    - Controls confounding variables

Best practices

1. Keep track of how results were produced
2. Avoid manual data manipulation
3. Document randomness
4. Interpretability
   - Understand the cause of a decision or predict model results
   - Story with compelling narrative
   - Link with reproducibility

[1] Molnar C. Interpretable Machine Learning. 2019.

Best practices

1. Keep track of how results were produced
2. Avoid manual data manipulation
3. Document randomness
4. Interpretability
5. Cite bibliography correctly

References

- A **citation** is the basic information required to **identify** and **locate** a specific publication

References

- Different styles but same underlying logic
  - *Book*: Author Name (Year). Title. Publisher.
  - *Journal Article*: Author Name. (Year) 'Article Title.' Journal Title, Volume Number, Issue Number, Page Numbers.
  - *Website*: Author Name. Date of Publication, 'Title of Page/Work.' Title of Website, Location

- **APA style**:
  - In text citations (author, date)

Reference

- Reference management tools
  - Easier to keep track
  - Change between styles
  - Search for reference online
  - Options:
    - EndNote
    - Mendeley
    - RefWorks

References

- Business context
  - Less strict
  - Simpler (hyperlink)
  - ==> information available and retrievable

Let's practice!

DATA COMMUNICATION CONCEPTS

Write precise and clear reports

DATA COMMUNICATION CONCEPTS

Hadrien Lacroix
Curriculum Manager

datacamp

Written report

A report must be **clear** and reproducible.

Write precise and clear reports

- Concise
- Precise
- Avoid misleading and confusion
- Meaningful message

[1] Nolan D, Stoudt S. Communicating with Data. OUP Oxford. 2021.

Empty phrases

- Contain no information
  - It is interesting to note that
  - The fact that
  - It should be pointed out that
  - It is well known that
  - It is obvious that

[1] Nolan D, Stoudt S. Communicating with Data. OUP Oxford. 2021.

Empty phrases

- Contain no information
- Distracting
- ==> should be removed

[1] Nolan D, Stoudt S. Communicating with Data. OUP Oxford. 2021.

Empty phrases

Negative ratings are associated with the words "delayed" and "shipping"

Another important point is the fact that negative ratings were associated with the words "delayed" and "shipping"

[1] Nolan D, Stoudt S. Communicating with Data. OUP Oxford. 2021.

Concrete nouns

- Write concrete nouns
- Avoid "this", "that", "it"
  - Adds cognitive load
  - Distracts them from insights

[1] Nolan D, Stoudt S. Communicating with Data. OUP Oxford. 2021.

Concrete nouns

This shows an accuracy of 80% when predicting customer churn.

The model shows an accuracy of 80% when predicting customer churn.

[1] Nolan D, Stoudt S. Communicating with Data. OUP Oxford. 2021.

More pronouns

- Active voice: emphasis on the author
- Passive voice: stuffy and hard to read
- Academic vs business context

[1] Nolan D, Stoudt S. Communicating with Data. OUP Oxford. 2021.

Redundant adjectives and adverbs

- Phrases that say the same thing twice
  - Introduce a new
  - Done previously
- Eliminate redundant adjective and adverbs

[1] Nolan D, Stoudt S. Communicating with Data. OUP Oxford. 2021.

Run-on sentences

- Two or more independent clauses connected incorrectly
  - There is a correlation between delayed shipping and customer rating, the shipping delay is the cause for negative review.
- Correction
  - Make two sentences
  - Use dependent clause

Let's practice!

DATA COMMUNICATION CONCEPTS

Case study: report on credit risk

DATA COMMUNICATION CONCEPTS

Hadrien Lacroix
Curriculum Manager

Credit risk

- Credit risk: probability of defaulting
- Loanme bank wants to predict if a customer is likely to default
- Raw data available
- Data Exploration Analysis
- Model training and evaluation

Audience

- Non-technical stakeholders
- Bank decision-makers

Story

- Background:
  - Increase in defaulting percentage over last 5 years.
  - Predicting which customers had a high probability of default.
- Insight: People with more unemployment periods tends to default more
- Insight: People with lower income tend to default more
- Climax: Possible to predict which people is more likely to default with an accuracy of 95%
- Next steps: Run a trial on a control population

Tech or non-tech

- Translate technical results

The right data

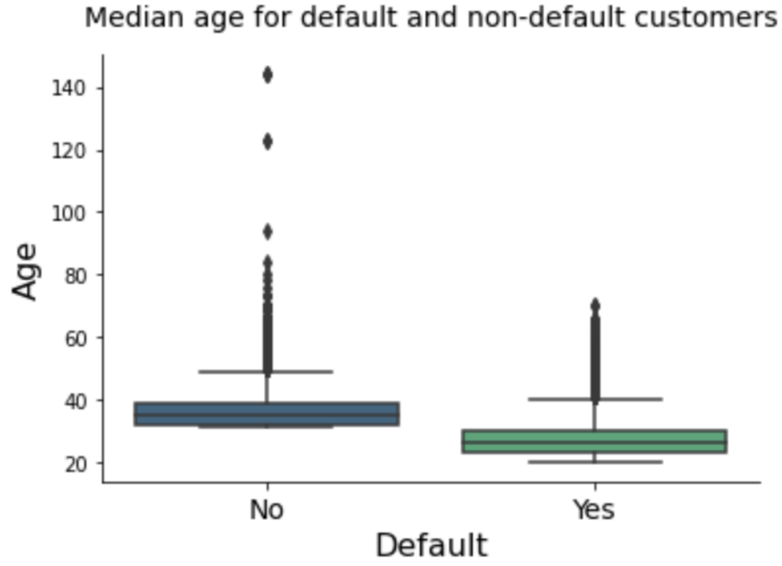
- Audience persona
  - **Role:** Financing Department Director
  - **Interest:** Decision on implementing an automated loan rejection system
  - **Appropriate data:**
    - Relationship between age or income and loan default
    - Percentage customer defaulting over the next months

Statistics

- Median age and income
- Percentage of change

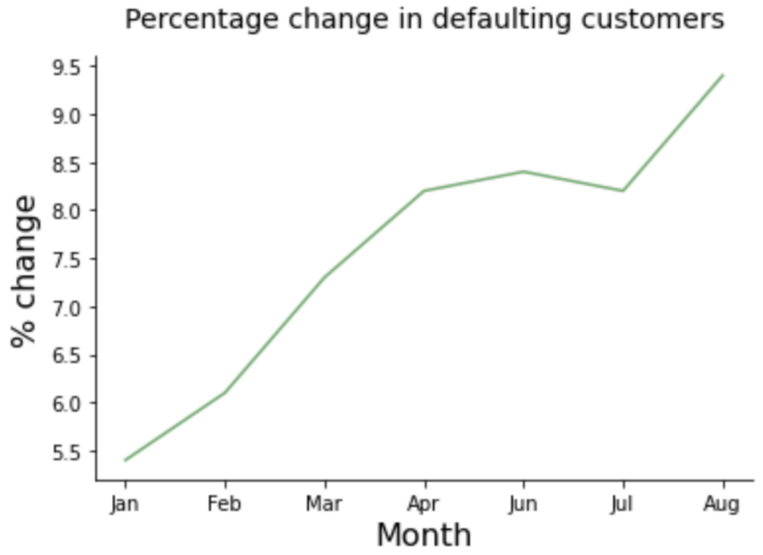
Visuals

- Boxplot with age vs. default condition



Visuals

- Boxplot with age vs. default condition
- Lineplot with % change defaulting customers



Correct format

- Who? **Financial Department director**
- Why? **Important decisions ahead**
- Content: **Key findings and recommendations**
- Channel: **Send the results before the meeting**

Report

- Written report
- Summary report or final report?

Report

- Summary report
- Informational report vs. analytical report?

Report

- Summary report
- Analytical report

Summary report structure

- Introduction
  - Purpose
  - Contextual information
  - Question of analysis
- Body
  - Data
  - Results: Key findings
- Conclusions
  - Restate question
  - Central insight
  - Add recommendations

Let's practice!

DATA COMMUNICATION CONCEPTS