Data Analytics and Visualisation

Vivek Katial

2024-02-05

Introduction

About Me

- Vivek Katial (vivek@gooddatainstitute.com)

About Me

- Vivek Katial (vivek@gooddatainstitute.com)
- Co-founder and Executive Director @ Good Data Institute

About Me

- Vivek Katial (vivek@gooddatainstitute.com)
- Co-founder and Executive Director @ Good Data Institute
- Data & Founding Team @ Multitudes

About Me

- Vivek Katial (vivek@gooddatainstitute.com)
- Co-founder and Executive Director @ Good Data Institute
- Data & Founding Team @ Multitudes
- PhD Candidate @ Unimelb (Optimisation on Quantum Computers)

About Me

- Vivek Katial (vivek@gooddatainstitute.com)
- Co-founder and Executive Director @ Good Data Institute
- Data & Founding Team @ Multitudes
- PhD Candidate @ Unimelb (Optimisation on Quantum Computers)
- Visiting PhD Researcher @ NASA Jet Propulsion lab

About Me

- Vivek Katial (vivek@gooddatainstitute.com)
- Co-founder and Executive Director @ Good Data Institute
- Data & Founding Team @ Multitudes
- PhD Candidate @ Unimelb (Optimisation on Quantum Computers)
- Visiting PhD Researcher @ NASA Jet Propulsion lab
- I love traveling and trying new types of food and meeting interesting people

Who is the Good Data Institute?

- We are nonprofit that empowers other nonprofits and social enterprises to leverage data for social impact.

Who is the Good Data Institute?

- We are nonprofit that empowers other nonprofits and social enterprises to leverage data for social impact.
- We build data capabilities through our volunteer community by working on data projects for the social sector

Who is the Good Data Institute?

- We are nonprofit that empowers other nonprofits and social enterprises to leverage data for social impact.
- We build data capabilities through our volunteer community by working on data projects for the social sector
- We are a global community of 150+ data nerds

Who is the Good Data Institute?

Our Impact

1. **75+** Data Projects
2. **50+** Nonprofit Partners
3. **150+** Data Nerds
4. **10+** Countries
5. **2500+** Volunteer Hours

Today's Agenda

1. Importance of Data Analytics and Visualisation
2. Key Tools for Nonprofits
3. Data Ethics
4. Real World Case Studies
5. Best Practices in Data Modeling and Visualisation
6. Q&A

Importance of Data Analytics and Visualisation

Why Data Matters

- motivation slide on why data matters for the social sector

What is Data Analytics and visualisation?

- slide about data analytics and visualisation and why they are

$$i\hbar\frac{\partial}{\partial t}\Psi = \hat{H}\Psi$$

Key Tools for Nonprofits

Tools and Technologies

- Free and Open Source Tools (R, Python libraries)
- Looker Studio
- Tableau Nonprofit Program
- Microsoft Power BI
- Custom Dashboards and Reports from Salesforce, etc.

How to Choose the Right Tool

- Where is your data stored already?
- What are your data visualization needs?
- What is your budget and technical capacity?

Pros and Cons of Different Tools

- **R and Python:** Highly customizable, but require coding skills and technical expertise
- **Looker Studio:** Easy to use, but limited customization
- **Tableau and Power BI:** Powerful features, but can be expensive
- **Custom Dashboards via ERPs:** Tailored to your needs, but require development resources

Data Ethics and Algorithmic Bias

What is Data Ethics and Algorithmic Bias?

- Data ethics refers to the moral and ethical implications of data collection, analysis, and use.

What is Data Ethics and Algorithmic Bias?

- Data ethics refers to the moral and ethical implications of data collection, analysis, and use.
- Algorithmic bias refers to the ability of algorithms to systematically and repeatedly produce outcomes that benefit one particular group over another

What is Data Ethics and Algorithmic Bias?

- Data ethics refers to the moral and ethical implications of data collection, analysis, and use.
- Algorithmic bias refers to the ability of algorithms to systematically and repeatedly produce outcomes that benefit one particular group over another
- Already many examples in society where algorithms have harmed marginalized groups

Trivial Example

- Predictions on the image of the Western bride included labels such as “bride”, “wedding”, “ceremony”

Trivial Example

- Predictions on the image of the Western bride included labels such as “bride”, “wedding”, “ceremony”
- For the woman wearing a traditional Indian wedding dress, the predicted labels were “costume”, “performing arts”, “event

More Harmful Example

[@Lum2016]

The Machine Learning Lifecycle

How to get started!

- Create some data principles

How to get started!

- Create some data principles
- This is a well-studied field. Get buy-in from leadership using existing research

How to get started!

- Create some data principles
- This is a well-studied field. Get buy-in from leadership using existing research
- Conduct a **bad actor exercise**

How to get started!

- Create some data principles
- This is a well-studied field. Get buy-in from leadership using existing research
- Conduct a **bad actor exercise**
- Make data labelling fun and diverse

How to get started!

- Create some data principles
- This is a well-studied field. Get buy-in from leadership using existing research
- Conduct a **bad actor exercise**
- Make data labelling fun and diverse
- Recognise that humans are the ones who create algorithms, so we also recognize the importance of the broader culture and environment we create and operate in.

How to get started!

- Create some data principles
- This is a well-studied field. Get buy-in from leadership using existing research
- Conduct a **bad actor exercise**
- Make data labelling fun and diverse
- Recognise that humans are the ones who create algorithms, so we also recognize the importance of the broader culture and environment we create and operate in.
- Commit to learning more!

Applying Data Analytics and Visualisation in the Social Sector

Case Study 1: Transforming Educational Insights in Colombia (Looker Studio)

- **Organization:** Red Inn Col
- **Challenge:** Tracking and reporting impact efficiently
- **Solution:** Implementation of Looker Studio for data visualization
- **Outcome:**

Case Study 2: Improving Model Performance for Drug Discovery (Ersilia) (MLOps)

- **Organization:** Ersilia
- **Challenge:** Improving model performance for drug discovery latency
- **Solution:** Rearchitecting the model pipeline using MLOps
- **Outcome:** Reduced latency by **90%**

Slide 6 - Best Practices in Data Visualization

- **Keep it simple:** Avoid clutter and focus on key messages
- **Be consistent:** Use uniform styles and colors
- **Tell a story:** Make data narrative and engaging
- **Accessibility:** Ensure visualizations are accessible to all audiences

Slide 7 - Q&A

Your Questions Answered

- Open floor for questions and interactive discussion

Slide 8 - Thank You

Stay Connected

- Contact information
- Links to additional resources
- Invitation to subscribe to our newsletter