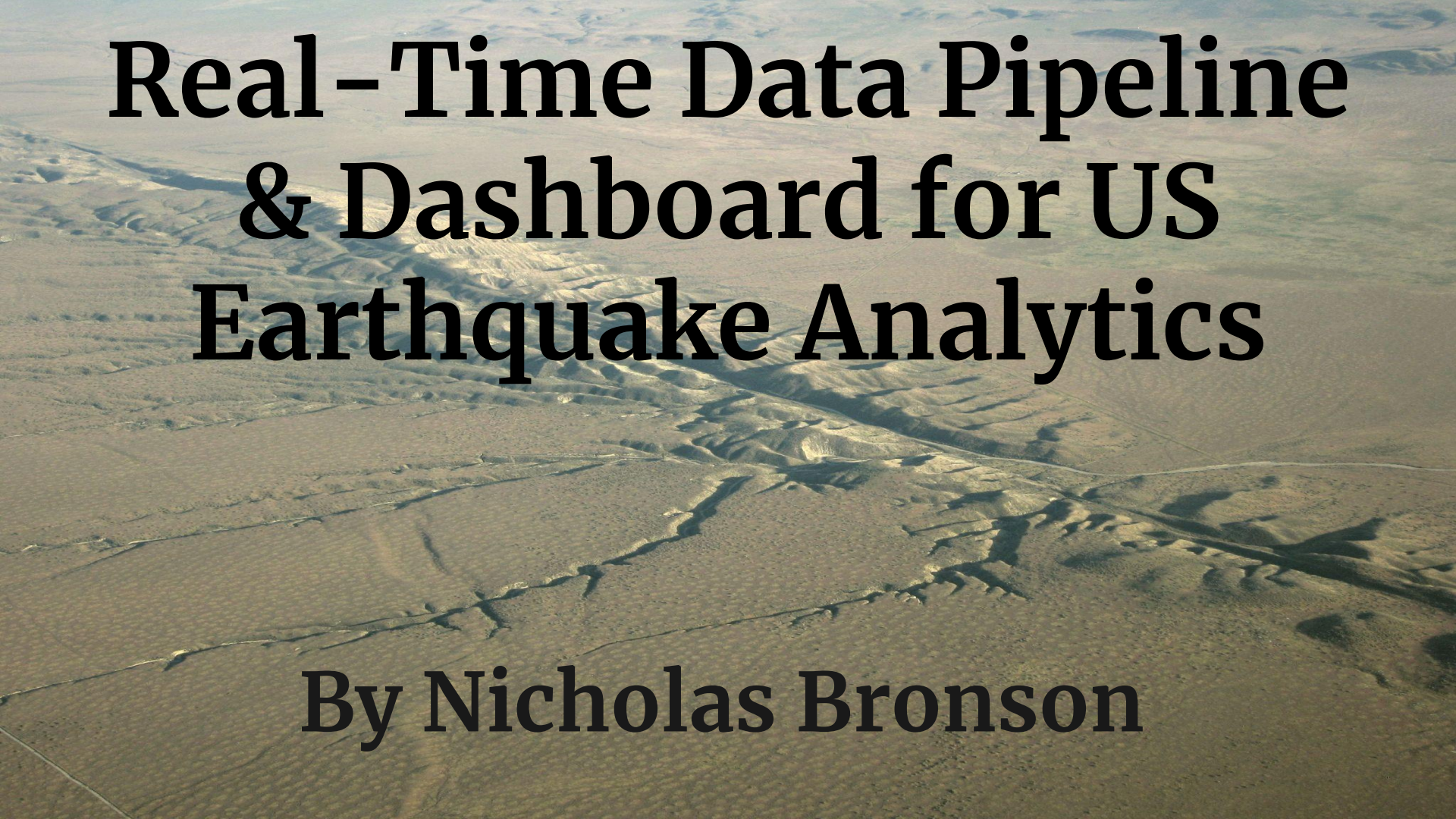


Real-Time Data Pipeline & Dashboard for US Earthquake Analytics

By Nicholas Bronson

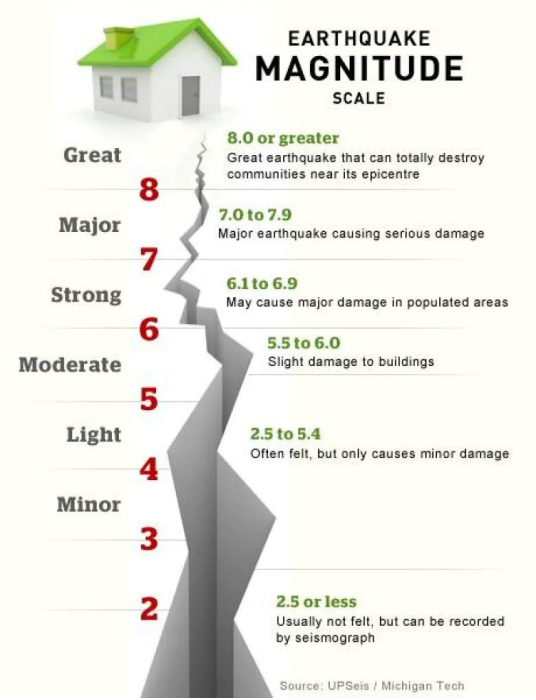
Agenda

1. Introduction & Questions
2. Data & Pipeline Setup
3. App Demonstration
4. Conclusions & Next Steps

1. Introduction & Problem

Introduction: Earthquake Basics

- Tectonic plates build tension and eventually slip past each other over an areas known as a fault line
- Earthquakes occur in waves, and vary drastically in strength, known as magnitude
- Earthquake magnitude is on a logarithmic scale, and an increase of 1 point can represent a dramatic difference in strength



Problem

Idea Catalyst: A CNN article asserted that the Blanco Fracture zone off the coast of Oregon has seen an uptick in earthquakes, some of which had a magnitude of 5.0+

Questions: Is the frequency of earthquakes occurring in Oregon and off its coast increasing? Can I provide a convenient way for those who are curious to investigate earthquake trends?

2. Data & Pipeline Setup

Data

- Over 870,943 rows of data representing all Recorded earthquakes since 2010
- Acquired the US Geological Survey Quakes API
  - 22 Features pulled, key features include: Magnitude, Depth, Magnitude Error, Latitude, Longitude, Depth, Depth Error, and ID
- Data is pulled weekly using an API call, processed, and then stored

Pipeline Setup

Data Pulled from USGS using an API call

Data ingested into the SQL database DB Browser

Data preprocessing and analysis using pySpark

PySpark

Further preprocessing using Pandas

Real- time data acquired using API, processed with minimal manual steps

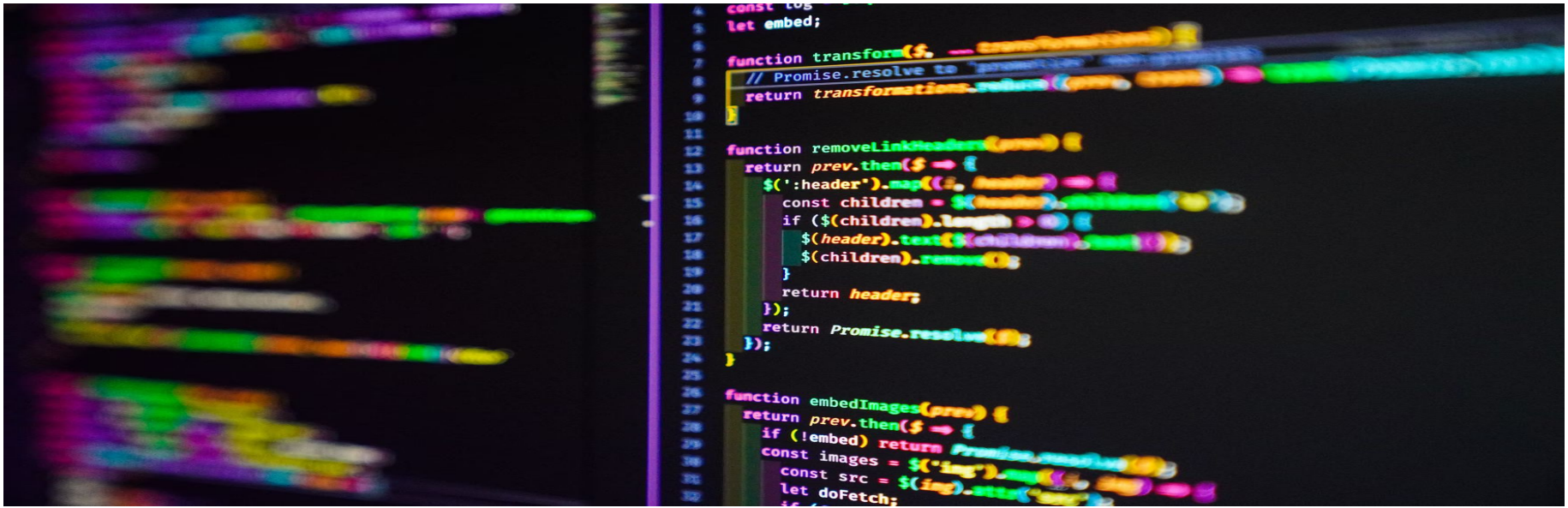
Data exported uploaded to GitHub and SQL

Streamlit Share page deployed, using data from GitHub

Streamlit

3. App Demonstration

https://share.streamlit.io/bronsonnh/streamlit_repo/main/nick-app.py



4. Conclusions & Next Steps

Conclusions

- There has been a **massive increase** in number earthquakes off the coast of Oregon. While this **does not appear to have caused any significant damage**, perhaps this trend is worth analyzing if it continues in subsequent years

- The vast majority of earthquakes in the database are **under the detectable level** – perhaps it would be worth focusing analysis on the relationship between lower level earthquakes and more impactful seismic events

Next Steps

- Investigate earthquake frequency and magnitude globally
- Post about seismic events as they occur, case studies on previous earthquakes
- Build a model that can predict number of earthquakes and strength of earthquakes in a given month


Thank you!