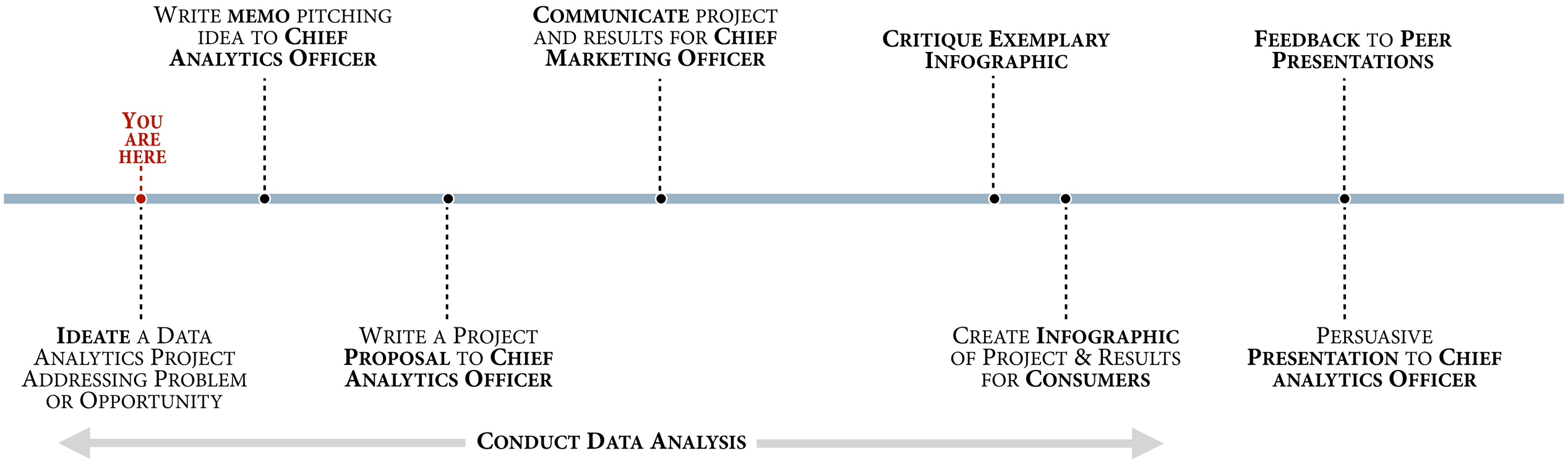
Storytelling With Data

Communication for applied analytics, challenges and scope

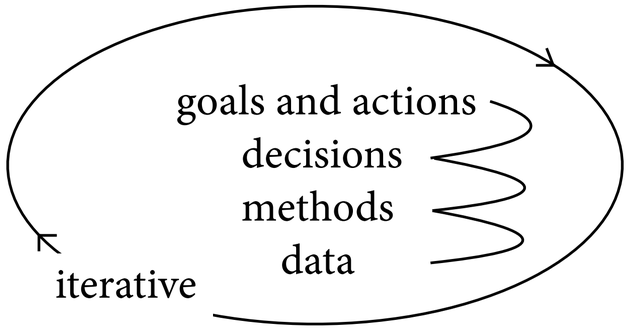
Scott Spencer | Columbia University

Conceptual project timeline



scoping a data analytics project

process



initial questions

What problem is to be solved?

Is the problem important?

Could an answer have impact?

Do data have a role in solving the problem?

Are the right data available?

In what contexts are the data generated?

Is the organization ready to tackle the problem and take actions from insights?

Spencer, Scott

DATUM : an abstraction of **a** real-world entity (person, object, or event). The terms *variable*, *feature*, and *attribute* are often used interchangeably to denote an individual abstraction.

Kelleher and Brendan

> “Data actually represent people — it represent us — because we human beings created this thing that we call data as a way to record our reality.”
>
> — Giorgia Lupi, *Information Designer*

DATA SET : consists of the data relating to a collection of entities, with each entity described in terms of a set of attributes. In its most basic form, a data set is organized in an $n \cdot m$ data matrix called the analytics record, where n is the number of entities (rows) and m is the number of attributes (columns).

Kelleher and Brendan

Data may be of different types, including nominal, ordinal, and numeric. These have subtypes as well.

Nominal types are *names* for categories, classes, or states of things.

Ordinal types are similar to nominal types, except it is possible to *rank or order* categories of an ordinal type.

Numeric types are *measurable* quantities we can represent using integer or real values. Numeric types can be measured on an *interval* scale or a *ratio* scale.

Kelleher and Brendan

scoping a data analytics project *for* Citi Bike

Friedman, Matthew. "***Citi Bike Racks Continue to Go Empty Just When Upper West Siders Need Them***." News. West Side Rag (blog), August 19, **2017**.

> "**Rebalancing** is one of the **biggest challenges** of any bike share system, especially in … New York where residents don't all work a traditional 9-5 schedule, and … people work in a variety of other neighborhoods."

— Simmons, Dani. Citi Bike spokeswoman.

Exercise 2.1: identify behaviors, events, data, and sources

Explore the availability of bikes and docking spots as **depending on users' patterns and behaviors, events and locations at particular times, other forms of transportation, and on weather**.

What **events** may be correlated with or cause empty or full bike docking stations? What potential user **behaviors** or **preferences** may lead to these events? From what analogous things could we draw **comparisons to provide context**?

How may these events and behaviors have been **measured** and **recorded**? What **data** are available? **Where** are it available? In what **form**?

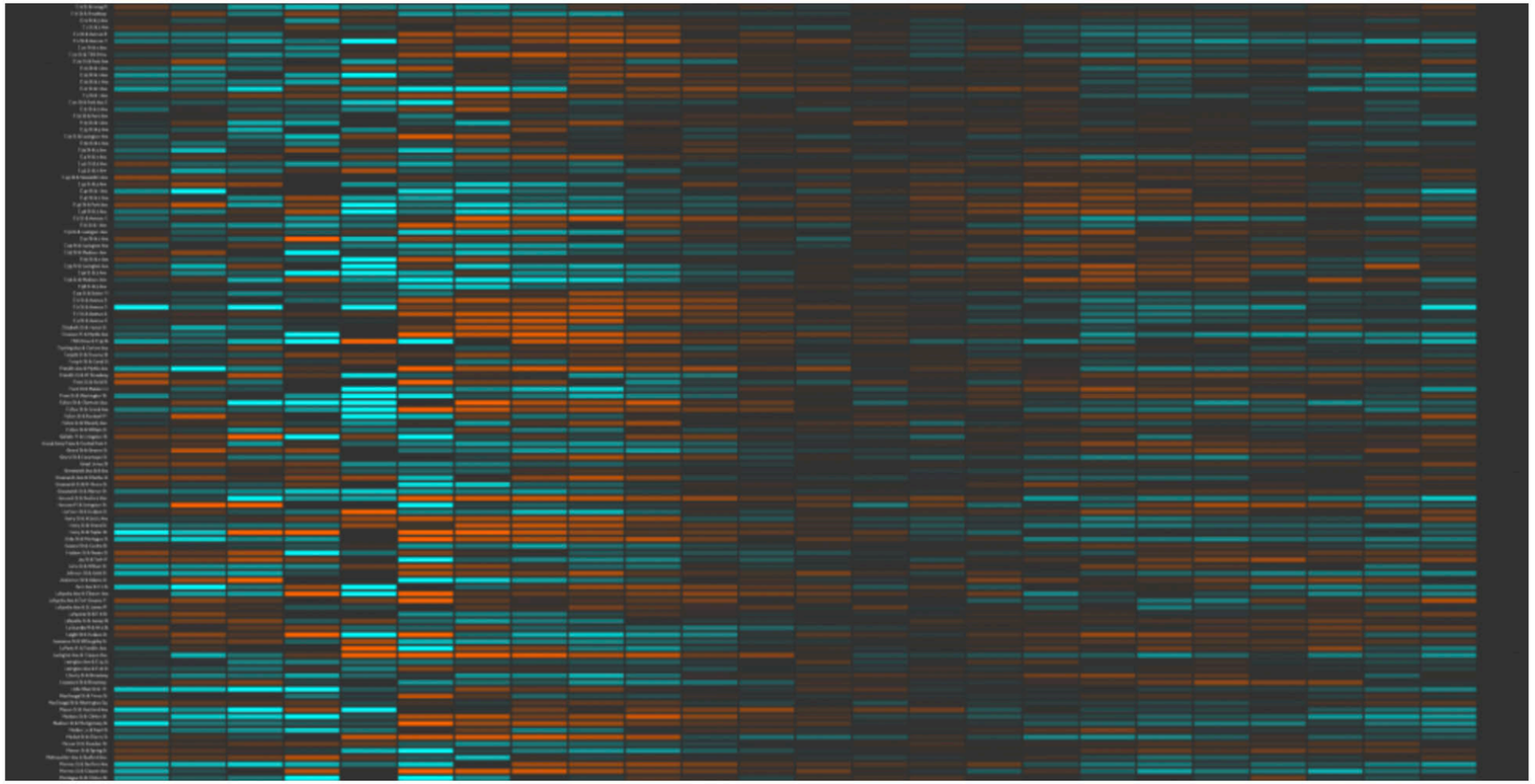
May these data be sufficient to find **insights** through analysis, useful for **decisions** and **goals**?

Get specific!

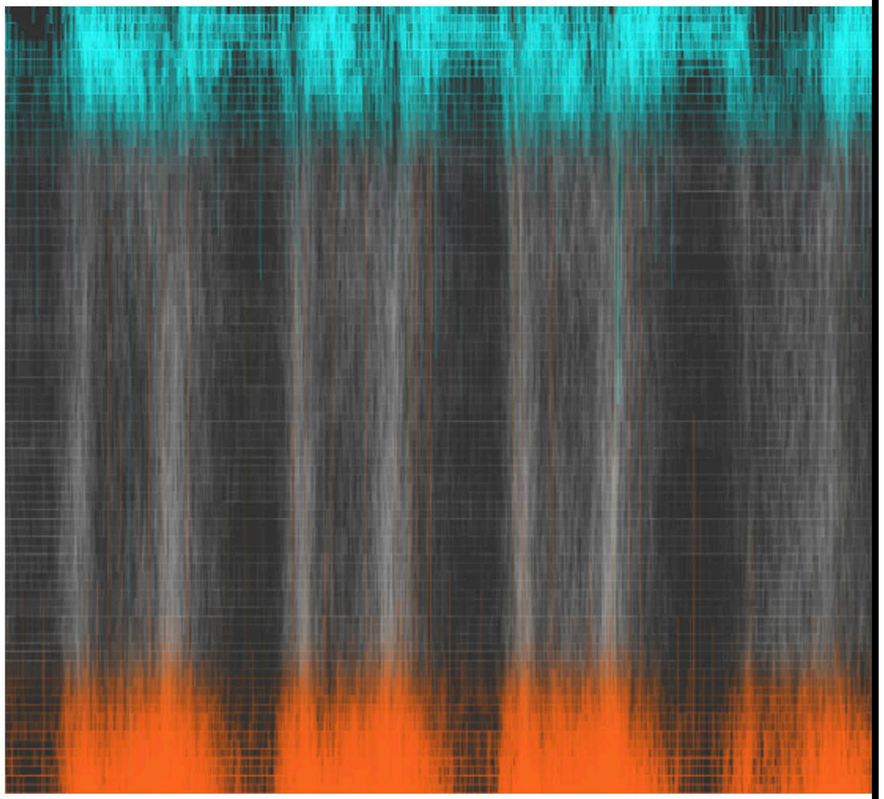
Project

CitiBike Rebalancing Study
An investigation into ways to rebalance CitiBike stations throughout New York City.

Browse by Initiative
Advanced Data Visualization



Close up of hourly balance matrix



Close up of stations balance chart

initial questions

What problem is to be solved?

Is the problem important?

Could an answer have impact?

Do data have a role in solving the problem?

Are the right data available?

Is the organization ready to tackle the problem and take actions from insights?

Exercise 2.1: identify behaviors, events, data, and sources

Data are recorded of each **bike** unlocked and docked, along with remaining **dock** capacities at the locations, dates, and times of each event: https://www.citibikenyc.com/system-data

Taxi pickup and drop-off locations and times: http://www.nyc.gov/html/tlc/html/about/trip_record_data.shtml

Subway lines entrance/exit locations: https://data.cityofnewyork.us/Transportation/Subway-Stations/arq3-7z49

Historical **weather**:
https://darksky.net/dev

Traffic data and more: http://www.nyc.gov/html/dot/html/about/datafeeds.shtml#realtime

discuss student project ideas

communicating data analytics

Audience for the memo, Chief Analytics Officer

Chief Analytics Officer | heads up a company's data analytics operations, transforming data into business value, and drives data-related business change.

—Zetlin, Minda. "What Is a Chief Analytics Officer? The Exec Who Turns Data into Decisions." CIO, November 2, 2017.

Minda Zetlin

Comparing project details described with varying lengths of an award-winning project

Post-project writeup | purpose and audience

"These **lessons** should be invaluable to the many **researchers and data scientists** who wish to partner with NGOs, governments, and other entities … working to use machine learning in the developing world."

In what ways are this audience and purpose similar to, and different from, the intended audience and purpose for our memos?

Improving Traffic Safety Through Video Analysis in Jakarta, Indonesia

João Caldeira*
Department of Physics
University of Chicago
jcaldeira@uchicago.edu

Alex Fout*
Statistics
Colorado State University
alex.fout@colostate.edu

Aniket Kesari*
Jurisprudence & Social Policy
University of California, Berkeley
akesari@berkeley.edu

Raesetje Sefala*
Machine Learning
University of the Witwatersrand
raesetje.sefala@students.wits.ac.za

Joseph Walsh
Center for Data Science and Public Policy
University of Chicago

Katy Dupre
Center for Data Science and Public Policy
University of Chicago

Muhammad Rizal Khaefi
Pulse Lab Jakarta
muhammad.khaefi@un.or.id

Setiaji
Jakarta Smart City
setiaji@jakarta.go.id

George Hodge
Pulse Lab Jakarta
george.hodge@un.or.id

Zakiya Aryana Pramestri
Pulse Lab Jakarta

Muhammad Adib Imtiyazi
Jakarta Smart City

Abstract

This project presents the results of a partnership with Jakarta Smart City (JSC) and United Nations Global Pulse Jakarta (PLJ) to create a video analysis pipeline for the purpose of improving traffic safety in Jakarta. The pipeline transforms raw traffic video footage into databases. By analyzing these patterns, the city of Jakarta will better understand how human behavior and built infrastructure contribute to traffic challenges and safety risks. The results of this work should also be broadly applicable to smart city initiatives around the globe as they improve urban planning and sustainability.

1 Introduction

The World Health Organization's *Global status report on road safety 2015* estimates that over 1.2 million people die each year in traffic accidents [1]. Nearly 2000 such fatalities occur annually in the city of Jakarta, Indonesia. Many of these deaths are preventable through effective city planning. Jakarta has experienced rapid population growth over the last 50 years, from roughly two million people in 1970 to more than 10 million today. With this growth comes a rise in vehicle ownership and congestion, leading to an increase in the number of traffic incidents

Improving Traffic Safety Through Video Analysis: Pulse Lab Jakarta.

Nearly 2,000 people die annually as a result of being involved in traffic-related accidents in Jakarta, Indonesia. The city government has invested resources in thousands of traffic cameras to help identify potential short-term (e.g. vendor carts in a hazardous location) and long-term (e.g. poorly engineered intersections) safety risks. However, manually analysing the available footage is an overwhelming task for the city's Transportation Agency. In support of the Jakarta Smart City initiative, our team hopes to build a video-processing pipeline to extract structured information from raw traffic footage. This information can be integrated with collision, weather, and other data in order to build models which can help public officials quickly identify and assess traffic risks with the goal of reducing traffic-related fatalities and severe injuries.

Caldeira, Joao

message first

Improving Traffic Safety Through Video Analysis: Pulse Lab Jakarta.

Nearly 2,000 people die annually as a result of being involved in traffic-related accidents in Jakarta, Indonesia. The city government has invested resources in thousands of traffic cameras to help identify potential short-term (e.g. vendor carts in a hazardous location) and long-term (e.g. poorly engineered intersections) safety risks. However, manually analysing the available footage is an overwhelming task for the city's Transportation Agency. In support of the Jakarta Smart City initiative, our team hopes to build a video-processing pipeline to extract structured information from raw traffic footage. This information can be integrated with collision, weather, and other data in order to build models which can help public officials quickly identify and assess traffic risks with the goal of reducing traffic-related fatalities and severe injuries.

Caldeira, Joao

Improving Traffic Safety Through Video Analysis: Pulse Lab Jakarta.

background context

Nearly 2,000 people die annually as a result of being involved in traffic-related accidents in Jakarta, Indonesia. The city government has invested resources in thousands of traffic cameras to help identify potential short-term (e.g. vendor carts in a hazardous location) and long-term (e.g. poorly engineered intersections) safety risks. However, manually analysing the available footage is an overwhelming task for the city's Transportation Agency. In support of the Jakarta Smart City initiative, our team hopes to build a video-processing pipeline to extract structured information from raw traffic footage. This information can be integrated with collision, weather, and other data in order to build models which can help public officials quickly identify and assess traffic risks with the goal of reducing traffic-related fatalities and severe injuries.

Caldeira, Joao

Improving Traffic Safety Through Video Analysis: Pulse Lab Jakarta.

goals, actions
origin of data

Nearly 2,000 people die annually as a result of being involved in traffic-related accidents in Jakarta, Indonesia. The city government has invested resources in thousands of traffic cameras to help identify potential short-term (e.g. vendor carts in a hazardous location) and long-term (e.g. poorly engineered intersections) safety risks. However, manually analysing the available footage is an overwhelming task for the city's Transportation Agency. In support of the Jakarta Smart City initiative, our team hopes to build a video-processing pipeline to extract structured information from raw traffic footage. This information can be integrated with collision, weather, and other data in order to build models which can help public officials quickly identify and assess traffic risks with the goal of reducing traffic-related fatalities and severe injuries.

Caldeira, Joao

Improving Traffic Safety Through Video Analysis: Pulse Lab Jakarta.

Nearly 2,000 people die annually as a result of being involved in traffic-related accidents in Jakarta, Indonesia. The city government has invested resources in thousands of traffic cameras to help identify potential short-term (e.g. vendor carts in a hazardous location) and long-term (e.g. poorly engineered intersections) safety risks.

problem

However, manually analysing the available footage is an overwhelming task for the city's Transportation Agency. In support of the Jakarta Smart City initiative, our team hopes to build a video-processing pipeline to extract structured information from raw traffic footage. This information can be integrated with collision, weather, and other data in order to build models which can help public officials quickly identify and assess traffic risks with the goal of reducing traffic-related fatalities and severe injuries.

Caldeira, Joao

Improving Traffic Safety Through Video Analysis: Pulse Lab Jakarta.

Nearly 2,000 people die annually as a result of being involved in traffic-related accidents in Jakarta, Indonesia. The city government has invested resources in thousands of traffic cameras to help identify potential short-term (e.g. vendor carts in a hazardous location) and long-term (e.g. poorly engineered intersections) safety risks. However, manually analysing the available footage is an overwhelming task for the city's Transportation Agency. In support of the Jakarta Smart City initiative, our team hopes to build a video-processing pipeline to extract structured information from raw traffic footage. This information can be integrated with collision, weather, and other data in order to build models which can help public officials quickly identify and assess traffic risks with the goal of reducing traffic-related fatalities and severe injuries.

method, data

Caldeira, Joao

Improving Traffic Safety Through Video Analysis: Pulse Lab Jakarta.

Nearly 2,000 people die annually as a result of being involved in traffic-related accidents in Jakarta, Indonesia. The city government has invested resources in thousands of traffic cameras to help identify potential short-term (e.g. vendor carts in a hazardous location) and long-term (e.g. poorly engineered intersections) safety risks. However, manually analysing the available footage is an overwhelming task for the city's Transportation Agency. In support of the Jakarta Smart City initiative, our team hopes to build a video-processing pipeline to extract structured information from raw traffic footage. This information can be integrated with collision, weather, and other data in order to build models which can help public officials quickly identify and assess traffic risks with the goal of reducing traffic-related fatalities and severe injuries.

impact, linked to goals, decisions

Caldeira, Joao

message first

Improving Traffic Safety Through Video Analysis: Pulse Lab Jakarta.

background context

goals, actions origin of data

problem

method, data

impact, linked to goals, decisions

Nearly 2,000 people die annually as a result of being involved in traffic-related accidents in Jakarta, Indonesia. The city government has invested resources in thousands of traffic cameras to help identify potential short-term (e.g. vendor carts in a hazardous location) and long-term (e.g. poorly engineered intersections) safety risks. However, manually analysing the available footage is an overwhelming task for the city's Transportation Agency. In support of the Jakarta Smart City initiative, our team hopes to build a video-processing pipeline to extract structured information from raw traffic footage. This information can be integrated with collision, weather, and other data in order to build models which can help public officials quickly identify and assess traffic risks with the goal of reducing traffic-related fatalities and severe injuries.

Caldeira, Joao

Improving Traffic Safety Through Video Analysis: Pulse Lab Jakarta.

Nearly 2,000 people die annually as a result of being involved in traffic-related accidents in Jakarta, Indonesia. The city government has invested resources in thousands of traffic cameras to help identify potential short-term (e.g. vendor carts in a hazardous location) and long-term (e.g. poorly engineered intersections) safety risks. However, manually analysing the available footage is an overwhelming task for the city's Transportation Agency. In support of the Jakarta Smart City initiative, our team hopes to build a video-processing pipeline to extract structured information from raw traffic footage. This information can be integrated with collision, weather, and other data in order to build models which can help public officials quickly identify and assess traffic risks with the goal of reducing traffic-related fatalities and severe injuries.

Caldeira, Joao

References

Caldeira, Joao, Alex Fout, Aniket Kesari, Raesetje Sefala, Joe Walsh, and Katy Dupre. "*Improving Traffic Safety Through Video Analysis: Pulse Lab Jakarta*." Data Science for Social Good (blog), 2018. https://dssg.uchicago.edu/project/improving-traffic-safety-through-video-analysis/.

Caldeira, Joao, Alex Fout, Aniket Kesari, Raesetje Sefala, Joseph Walsh, Katy Dupre, Muhammad Rizal Khaefi, et al. "*Improving Traffic Safety Through Video Analysis in Jakarta, Indonesia*." In Nd Conference on Neural Information Processing Systems NeurIPS, 1–5, 2018.

Friedman, Matthew. "*Citi Bike Racks Continue to Go Empty Just When Upper West Siders Need Them*." News. West Side Rag (blog), August 19, 2017. https://www.westsiderag.com/2017/08/19/citi-bike-racks-continue-to-go-empty-just-when-upper-west-siders-need-them.

Kelleher, John D, and Brendan Tierney. *Data Science*. MIT Press, 2018.

Saldarriaga, Juan Francisco. "*CitiBike Rebalancing Study.*" Spatial Information Design Lab, Columbia University, 2013. http://spatialinformationdesignlab.org/projects/citibike-rebalancing-study.

Zetlin, Minda. "*What Is a Chief Analytics Officer? The Exec Who Turns Data into Decisions*." CIO, November 2, 2017.