Introduction to AI

Moinul Zaber, Ph.D.
zaber@du.ac.bd
University of Dhaka
United Nations University
Data and Design Lab, dndlab.org

Why traditional algorithms may not be adequate ?

- Imagine we want to write a program to recognize a 3D object from a novel view point in a new lighting condition in a cluttered scene
  - How is it done in our brain ?
    - We do not know
  - So we do not know what program to write
    - Even if we knew how to, it would be a very complicated program

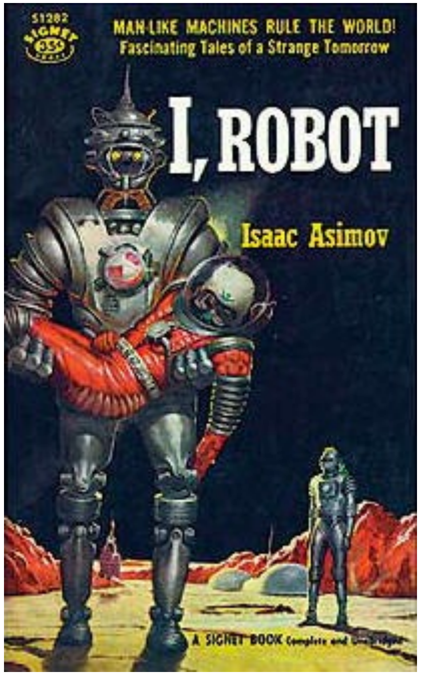
Image from : https://www.researchgate.net/publication/224167016_Stereo-Assist_Top-down_Stereo_for_Driver_Assistance_Systems

AI can be more efficient for many tasks

- Instead of writing a program for each specific task , we can collect a lot of examples that specify correct (incorrect) output of a given input
- AI algorithm then takes these examples and produces program that does the job
  - Program should work for new cases
  - If the data changes the program can change too by training on the new data
  - Massive amounts of computations are cheaper than writing task-specific program

AI – in the fiction

- Three laws of Robotics (Isaac Asimov)
  - *A robot may not injure a human being or, through inaction, allow a human being to come to harm.*
  - *A robot must obey orders given it by human beings, except where such orders would conflict with the First Law.*
  - *A robot must protect its own existence as long as such protection does not conflict with the First or Second Law.*



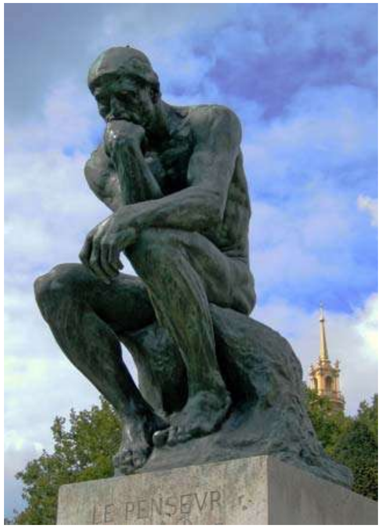
We are far from achieving “artificial super intelligence”

- Nor we have the capacity to build “General AI”
  - Refers to AI that is human like
  - Still decades or more to even manifest
- What we now have is popularly termed as “Narrow AI”
  - Systems that are intelligent not because they imitate human intelligence but because they can carry out tasks that would otherwise require human intelligence, time and effort to an unsustainable extent

This talk will only focus on issues related to Narrow AI

AI tries to model how humans THINK

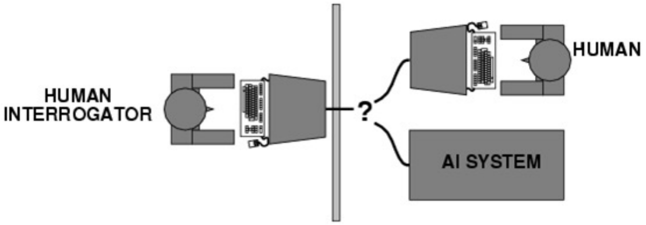
- Modeling exactly how humans **actually think**
  - Cognitive models of human reasoning
- Modeling exactly how humans **actually act**
  - Models of (what humans do, not how they think)
- Modeling how ideal agents **should think**
  - Models of “rational” thought (formal logic)
- Modeling how ideal agents **should act**
  - Rational actions but not necessarily formal rational reasoning
- Modern AI focuses on the last definition
  - Success is judged by how well the agent perform
  - Inspired by cognitive & neuroscience (how people think)



Augute Rodin: The Thinker, 1904,Bronze Sculpture ,PC: Shawn McCullars

Can Machines THINK ?: Turing Test

- Can machines behave intelligently ?
  - Imitation Game- Operational test for intelligent behavior
  - Allan Turing (1950) in “Computing machinery and intelligence” proposed this test
  - Here the interrogator, is given the task of trying to determine which player is a computer and which is a human.
  - The interrogator is limited to using the responses to written questions to make the determination



Can machine exhibit intelligent behaviour equivalent to, or indistinguishable from, that of a human?

AI needs an Inter Disciplinary approach

Philosophy	Logic, methods of reasoning, mind as physical system, foundations of learning, language, rationality.
Mathematics	Formal representation and proof, algorithms, computation, (un)decidability, (in)tractability, probability.
Economics	Utility, decision theory
Neuroscience	Neurons as information processing units.
Cognitive Science	how do people behave, perceive, process information, represent knowledge
Computer engineering	building fast computers
Control theory	design systems that maximize an objective function over time
Linguistics	knowledge representation, grammar

The state-of-the-art is still primitive

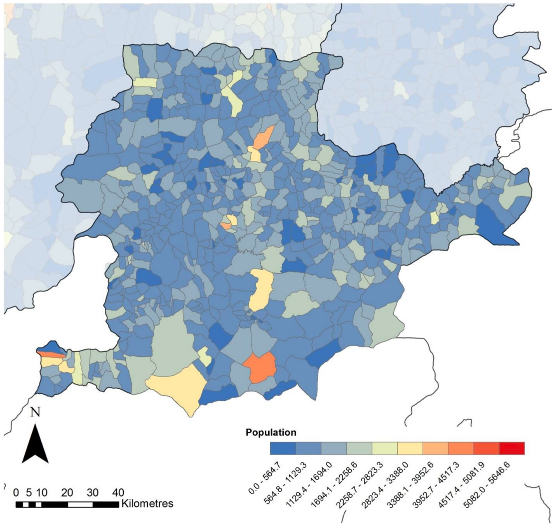
- Speech synthesis, recognition and understanding
  - Works for usefully limited vocabulary applications
  - Unconstrained speech understanding is still too hard
- Computer vision
  - Works well for constrained problems (hand-written zip-codes)
  - Understanding real-world, natural scenes is still too hard
- Learning
  - Adaptive systems are used in many applications but they have their limits
- Planning and Reasoning
  - Only works for constrained problems: e.g., playing chess
  - Real-world is too complex for general systems
- Overall:
  - Many components of intelligent systems are “doable”
  - List of many interesting research problems is very long

AI is used for Social Good

- Crisis response
- Economic empowerment
- Educational challenges
- Environmental challenges
- Equality and inclusion
- Health and hunger
- Information verification and validation
- Infrastructure management
- Public and social sector management
- Security and justice

Afghanistan Census (UNFPA)

- Last census conducted in 1979
  - Insecurity prevented attempts
- In 2017 UNFPA with other organizations started a hybrid census to generate spatially disaggregated population estimates
  - Prediction of population estimates in un-surveyed areas –modelling based on ground surveys and satellite imaging



Source : https://www.unfpa.org/resources/new-methodology-hybrid-census-generate-spatially-disaggregated-population-estimates

Forecasting Flood (Google)

- Depict a flood simulation of a river in Hyderabad, India. The left side uses publicly available data while the right side uses Google data and technology. Right hand side models contain higher resolution, accuracy, and up-to-date information.

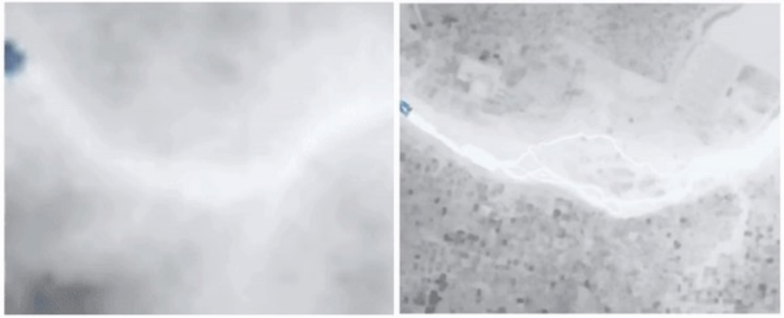
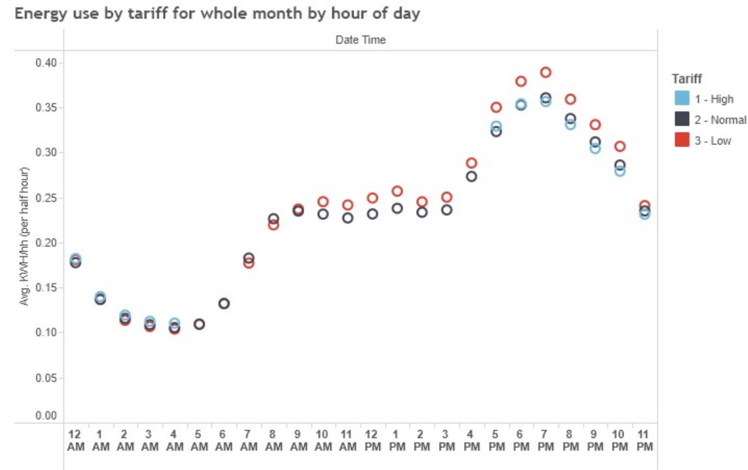


Image: https://www.blog.google/products/search/helping-keep-people-safe-ai-enabled-flood-forecasting/

Data from sensors come in various formats

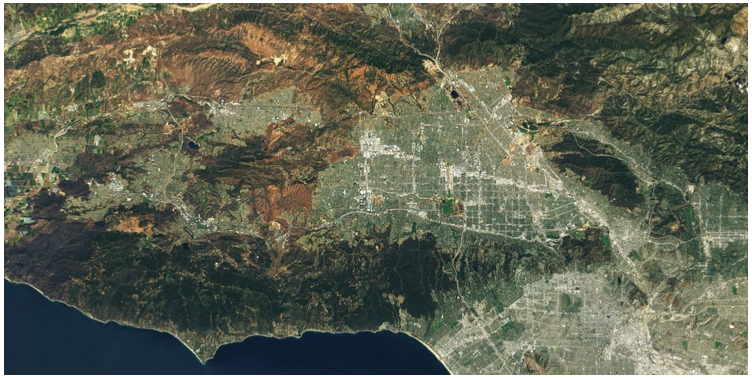
- Most sensors and smart meters provide time-series data mostly available as tabular data and can be stored as csv, Excel Spreadsheets or in a database
- Traffic camera record and store video data as timestamped image frames and is available in different formats.
- Satellites capture very large images, often with geographic data embedded and is available as GeoTIFF or other scientific data formats (h5 and others)
- Sources:
    - Time series data: [SmartMeter Energy Consumption Data in London Households]
    - Video Frames: Wikipedia ([link])
    - Satellite image: Los Angeles, Landsat band 2,3 & 4. ([nasa.gov])

Time series data



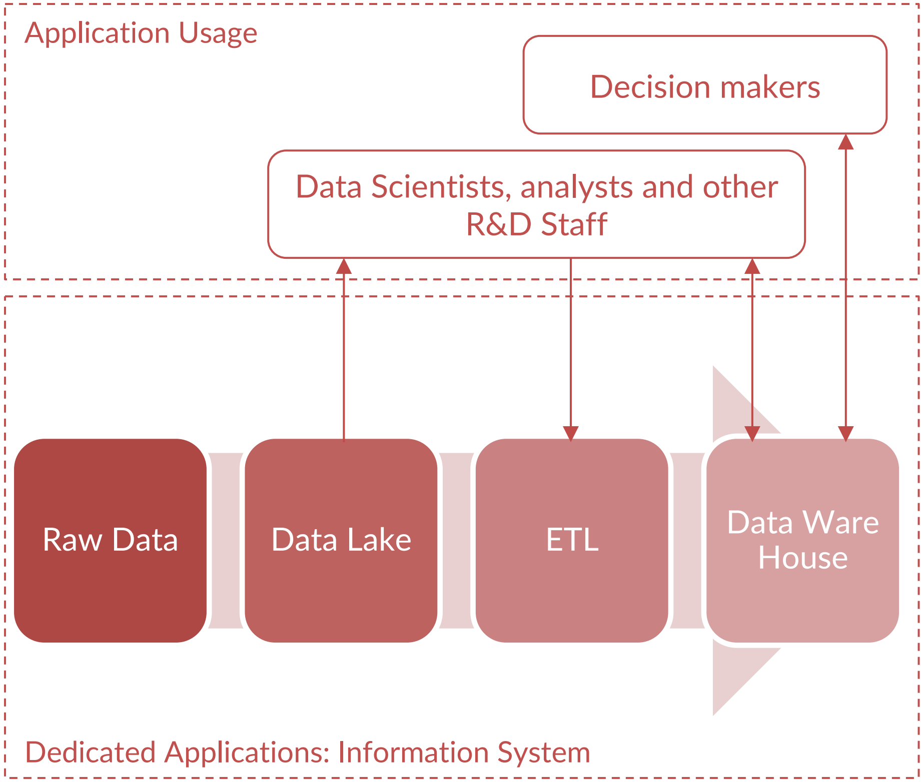
Video Frames

Landsat Image



Raw data is stored and transformed into structured data by R&D staff who use it to extract insights that is used by decision makers

- Raw but unstructured data is stored in the data lake
- R&D staff use the raw data to create Extract-Load-Transform (ETL) pipelines, which involves
  - preprocessing: data cleaning, validating and imputation
  - Exploratory analysis based on requirements
  - Transforming data into structured dataset to be stored in the warehouse for later use.
- The structured data is later used for analytics, reporting and creating data products
- Designated users access these from edge devices. The computing power of the devices depends on user types.



Non traditional data sources can either complement or can replace the tradition data by provide high frequency data at higher spatial resolution

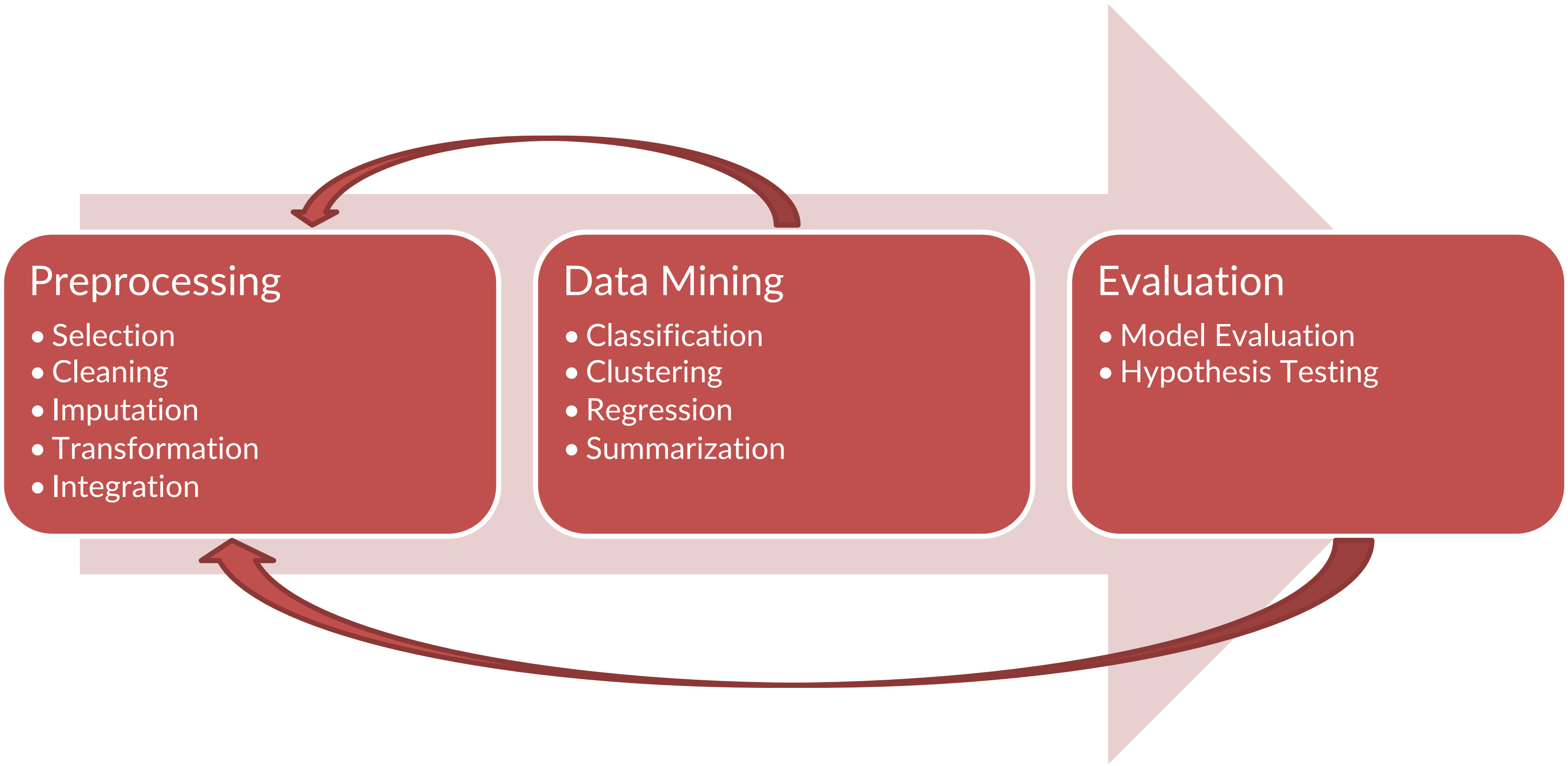
Example of non-traditional data for Statistical Applications

Data	Statistical application	References
Mobile CDR	Estimating poverty and wealth	Blumenstock et al. (2015)
	Human mobility and socioeconomic levels	Frias-Martinez et al. (2012)
	Socioeconomic status	Gutierrez et al. (2013)
	Creditworthiness of the unbanked	Kumar and Mohta (2012)
	Human mobility after disasters	Lu et al. (2016); Wilson et al. (2016);Lu etal. (2012)
Satellite image and remote sensing	Identifying the poor	Elvidge et al. (2009); Jean et al. (2016)
	Urban poverty	Kohli etal. (2012)
Mobile CDR	Proxy indicator for food expenditure	Decuyper et al. (2014)
Satellite image and remote sensing	Severity and extent of drought conditions	Berhan et al. (2011); Tucker & Choudhury (1987); Henricksen, & Durkin (1986);
	Developing vegetation health indices	Kogen et al. (2011)
Social media / online data	Constructing consumer price index	Cavallo and Rogobon (2016)
Satellite image and remote sensing	Identifying Poverty	Jean el al. (2016)
	Identifying slums	Kohli etal. (2012)
	Land cover/land use changes	Tso & Mather (2001); Lu & Weng, (2007); Thomas et al. (2011)

Source: Lokanathan et al. (2017) (pdf).

To make data useful for AI it often requires going back and forth till some of value is found

- Often times you will run into problems that require you to add more data (go back to preprocessing)
  - Data might make sense if some merged with another data
  - Or maybe your models/hypothesis failed and now you are trying new approaches



Preprocessing

- Selection
  - Understanding the data and select pertinent data

- Cleaning
  - Find any missing data
  - Validate data columns (e.g. working with human input often times misspell)
  - Check if data is balanced (e.g. working on a classification problem or varying observations)

- Imputation
  - Replace missing data using any statistical measures (e.g. mean, median or mode)
  - Use clues from other columns (e.g. find rows that are similar and copy values)

- Transformation
  - Understand data distribution of the columns
  - Normalize data that are on different scale
  - Create new columns from existing columns

- Integration
  - Merge/join datasets from different sources (e.g. Joining electricity consumption with weather data)

Data Mining

- Summarization
  - Aggregate data and use statistical measures
  - Create visualization that show trends, summaries, etc.

- Clustering
  - Group data by similarity
  - Understand common patterns in groups

- Modelling
  - Prepare features
  - Create empirical model to understand effects of independent variables (e.g. linear regression)
  - Use machine learning or Deep learning models
    - Categorization
    - Prediction
    - Computer vision
    - Natural Language processing

Evaluation

- Model Evaluation
  - Checking model fit (e.g. cross validation, Overfitting and under fitting)

- Hypothesis Testing
  - Evaluate model parameters and summaries
  - run statistical tests

Imputation Example: Working crime incident dataset from San Francisco

- San Francisco police department has provided data on **all reported incident from 2016**
  - Newer data can be found from civichub.us

A missing district value, thankfully we have location data that can be used to impute this cell

Different crime category

Description

Different districts

Outcome

Readable address

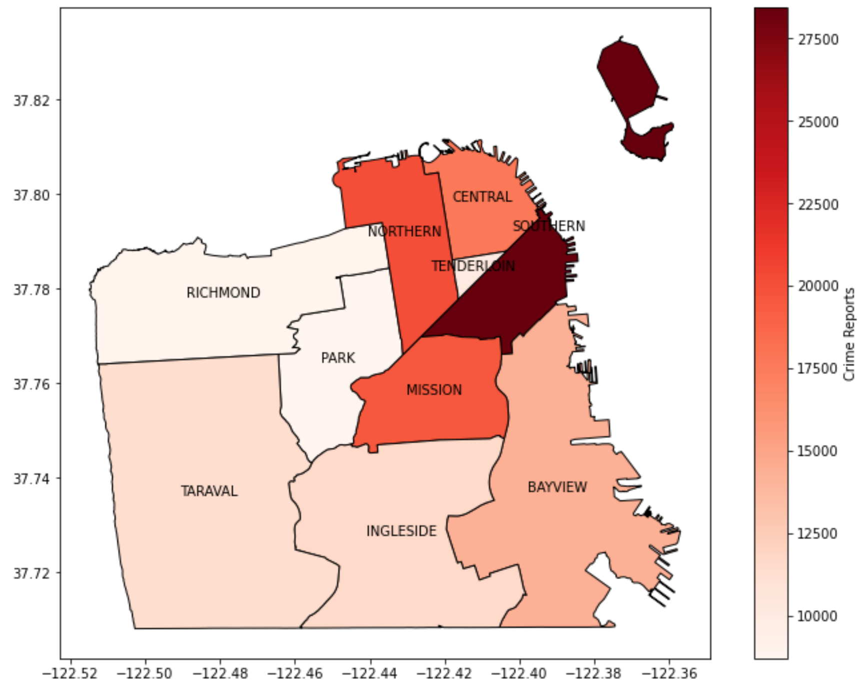
Coordinates

	IncidntNum	Category	Descript	DayOfWeek	Date	Time	PdDistrict	Resolution	Address	X	Y	PdId
112849	166018551	LARCENY/THEFT	GRAND THEFT FROM UNLOCKED AUTO	Monday	01/18/2016 12:00:00 AM	22:40	INGLESIDE	NONE	PRECITA AV / HARRISON ST	-122.411	37.746	16601855106224
112850	166018567	LARCENY/THEFT	GRAND THEFT FROM LOCKED AUTO	Tuesday	01/19/2016 12:00:00 AM	08:55	INGLESIDE	NONE	700 Block of CORTLAND AV	-122.414	37.738	16601856706244
112851	166018573	LARCENY/THEFT	GRAND THEFT FROM LOCKED AUTO	Sunday	01/17/2016 12:00:00 AM	23:54	NaN	NONE	100 Block of VELASCO AV	-122.413	37.708	16601857306244
112852	166018589	LARCENY/THEFT	GRAND THEFT FROM LOCKED AUTO	Thursday	01/14/2016 12:00:00 AM	17:00	INGLESIDE	NONE	1400 Block of CORTLAND AV	-122.409	37.739	16601858906244

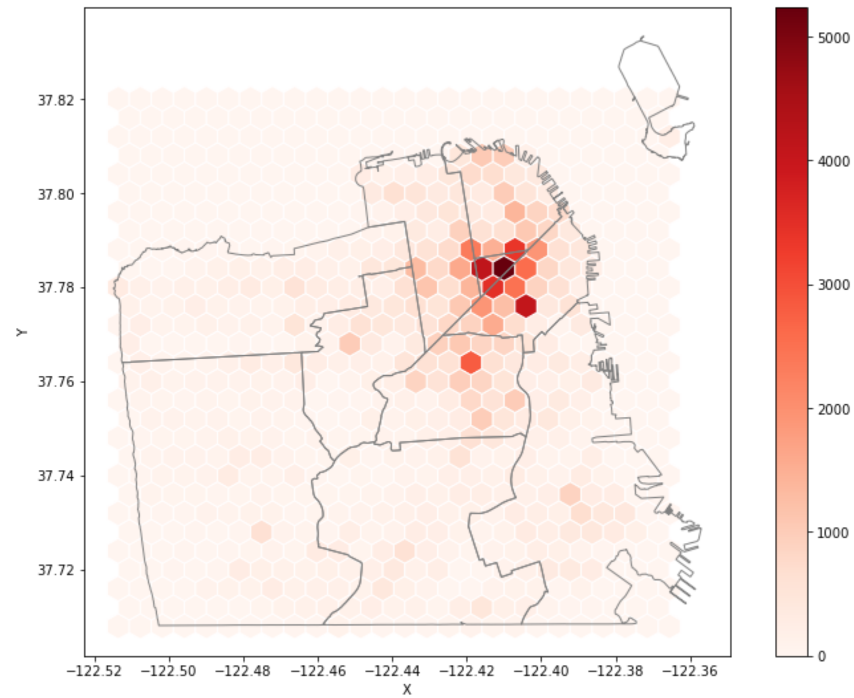
5 of 150,500 rows in the dataset

Once preprocesses, dataset can be summarized for insights, for example, where crime is most frequent

Crime hotspots by districts



Crime hotspots by hex binning



Predicting Electricity consumption using smart meter data of London households, weather and holidays

- UK Power Networks led Low Carbon London project
  - between November 2011 and February 2014
  - 5,567 London Households that took part
- And has provided the open access to the [dataset]
- Imported weather data from Darksky API
- Imported list of holidays from holiday calendar

Date / Household ID / Total energy consumed

day	LCLid	energy_sum
2012-10-12	MAC000002	7.098
2012-10-13	MAC000002	11.087
2012-10-13	MAC000002	11.087
2012-10-13	MAC000002	11.087
2012-10-13	MAC000002	11.087

Snapshot of the energy dataset with household ID

Max temperature and time / Other Weather data

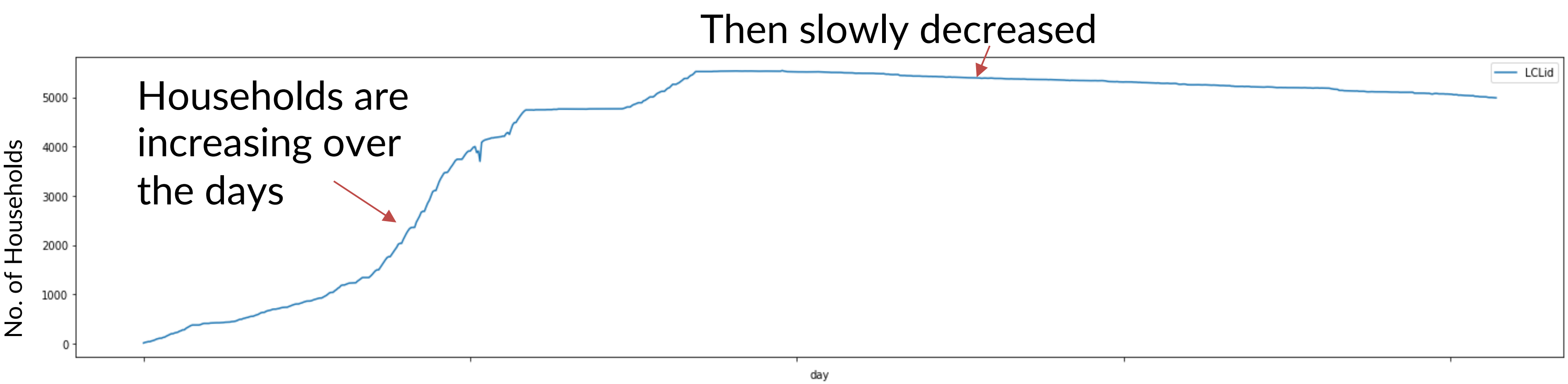
temperatureMax	temperatureMaxTime	windBearing	icon	dewPoint
11.96	11-11-11 23:00	123	fog	9.4
8.59	11-12-11 14:00	198	partly-cloudy-day	4.49
10.33	27-12-11 2:00	225	partly-cloudy-day	5.47
8.07	02-12-11 23:00	232	wind	3.69

Snapshot of the weather data with time

Participants increased and then steadily decreased, why?

- Some reasons might be:
  - Households were asked to participate and they responded over time
  - Some households opted out over time

- To overcome the issue, we can use average consumption instead



Dataset has average household consumption from 829 days with a mean of 10.49 KWH and standard deviation of 1.9 KWH

- Get an understanding of the Average Energy data
    - Mean = 10.49 KWH
    - Standard deviation = 1.90 KWH
    - Minimum and maximum = 0.21 KWH, 15.9 KWH respectively
    - Different Quartiles (Q1, Q2, Q3) = 8.67, 10.52, 12.00 KWH respectively

	energy_sum	LCLid	avg_energy
count	829.000000	829.000000	829.000000
mean	43535.325676	4234.539204	10.491862
std	20550.594031	1789.994799	1.902513
min	90.385000	13.000000	0.211766
25%	34665.436003	4084.000000	8.676955
50%	46641.160997	5138.000000	10.516983
75%	59755.616996	5369.000000	12.000690
max	84156.135002	5541.000000	15.964434

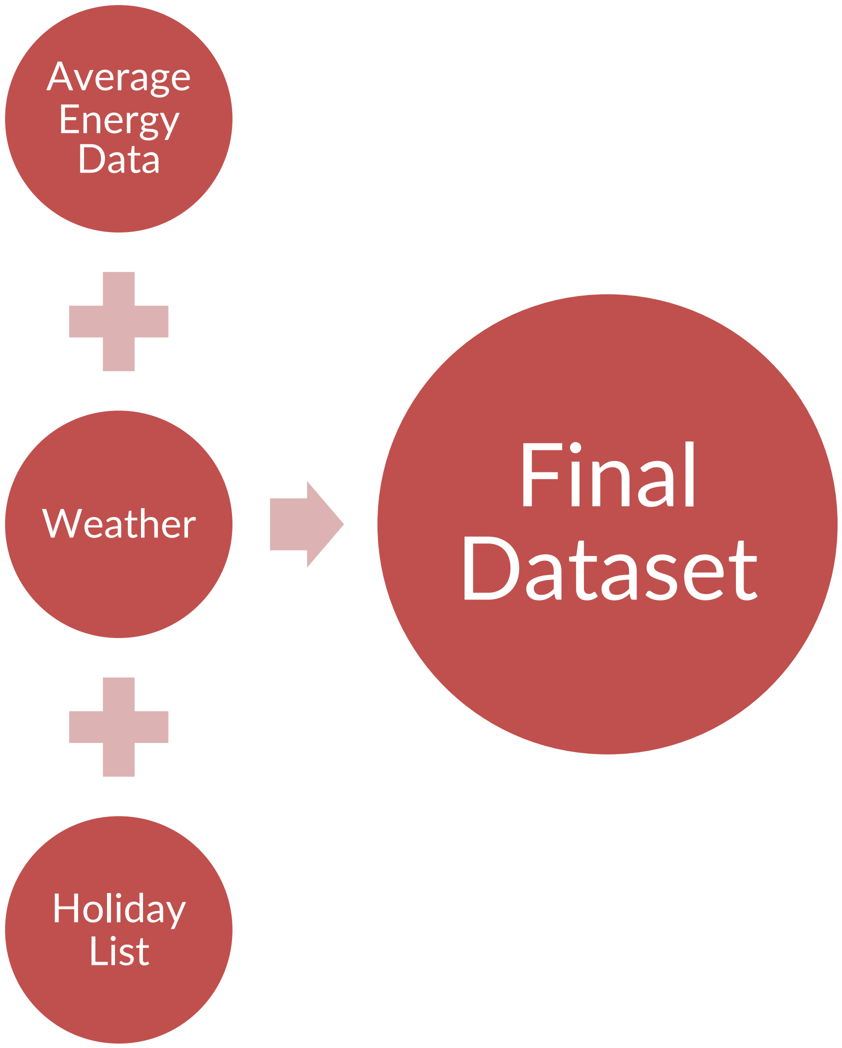
Description of energy dataset

	temperatureMax	windBearing	dewPoint	cloudCover	windSpeed	pressure
count	882.000000	882.000000	882.000000	881.000000	882.000000	882.000
mean	13.660113	195.702948	6.530034	0.477605	3.581803	1014.12
std	6.182744	89.340783	4.830875	0.193514	1.694007	11.0730
min	-0.060000	0.000000	-7.840000	0.000000	0.200000	979.250
25%	9.502500	120.500000	3.180000	0.350000	2.370000	1007.43
50%	12.625000	219.000000	6.380000	0.470000	3.440000	1014.61
75%	17.920000	255.000000	10.057500	0.600000	4.577500	1021.75
max	32.400000	359.000000	17.770000	1.000000	9.960000	1040.92

Description of the weather data

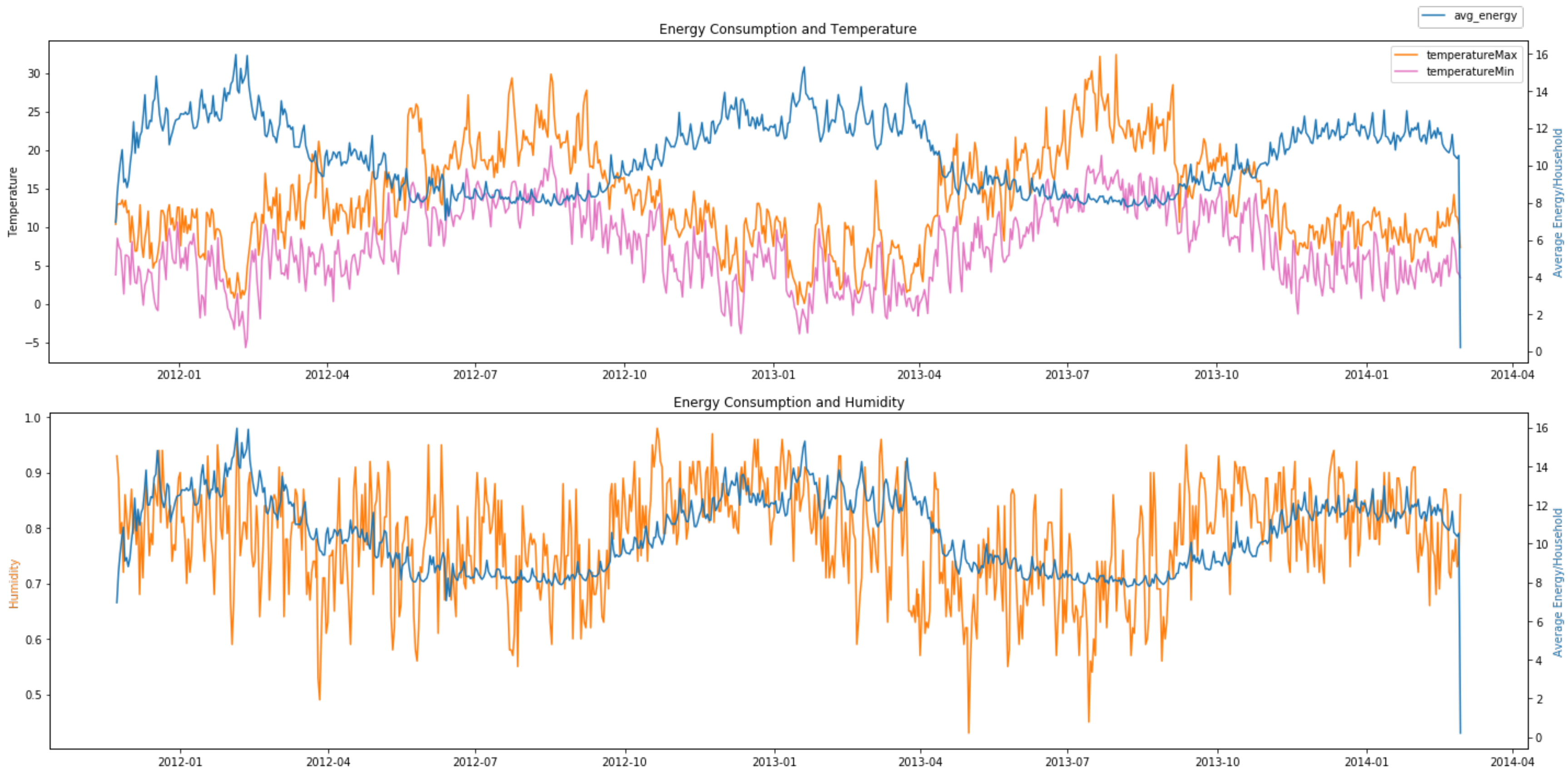
Integration: Combine datasets

- Data Columns
  - Date
  - Average energy (numeric)
  - Weather columns (numeric)
  - Holiday (Boolean, 0 or 1)



Observe relationship: Temperature and Humidity

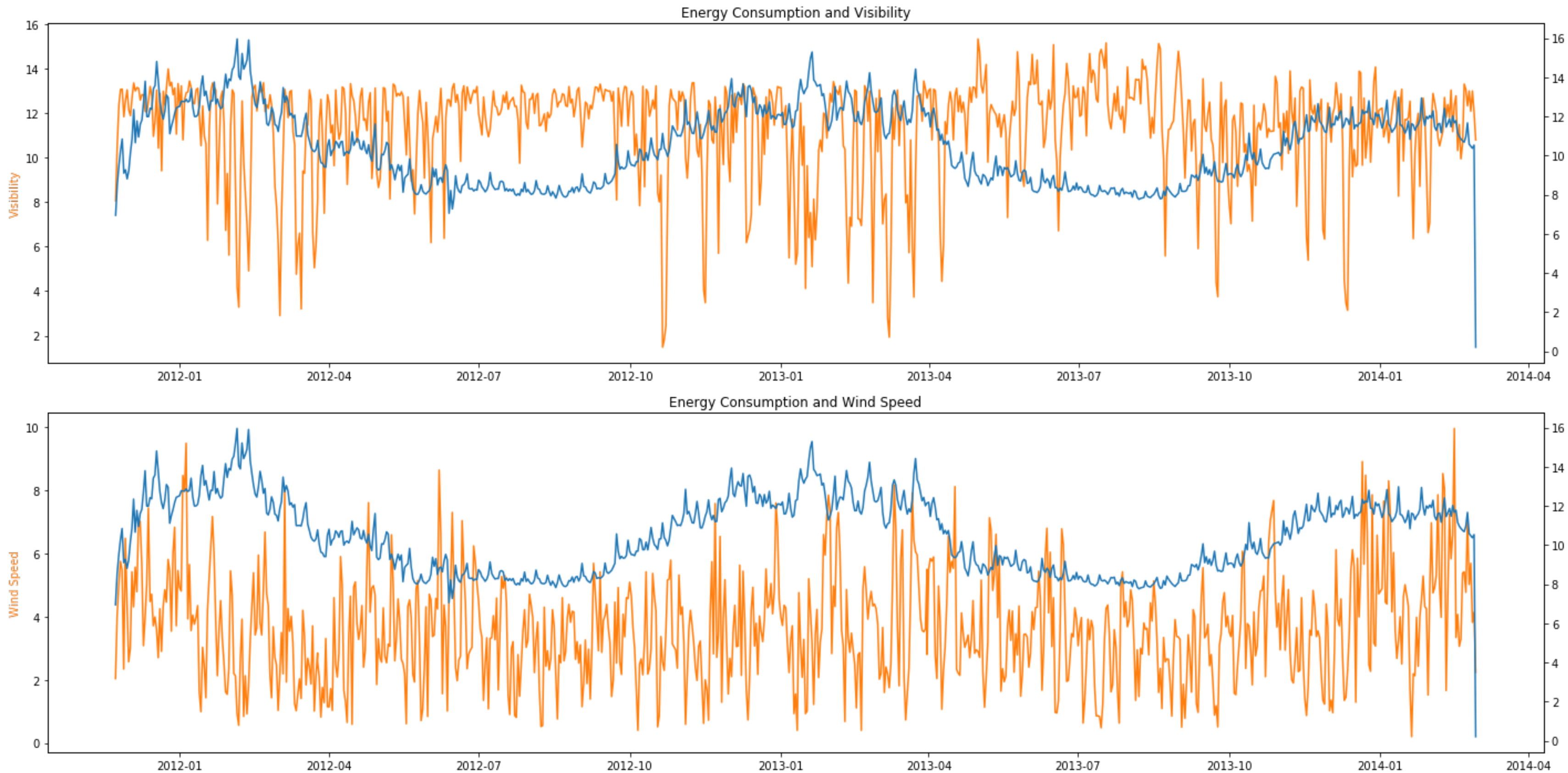
- We observe that consumption has an inverse relationship with temperature
- But varies closely with humidity



N.B. Y-axes are different to align data visualization

Observe relationship: Visibility and wind speed

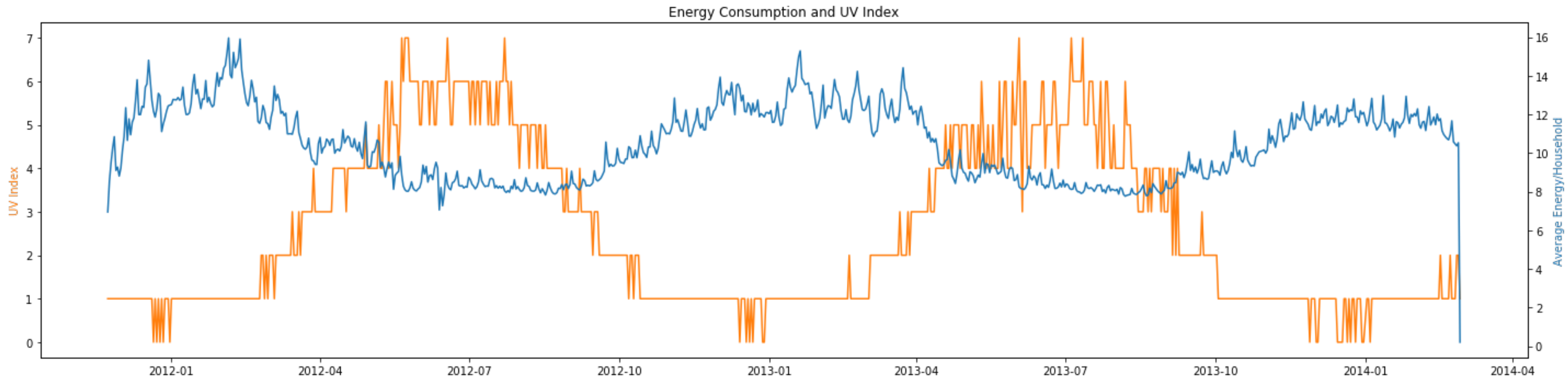
- We observe that consumption may not have any close relationship with visibility and wind speed

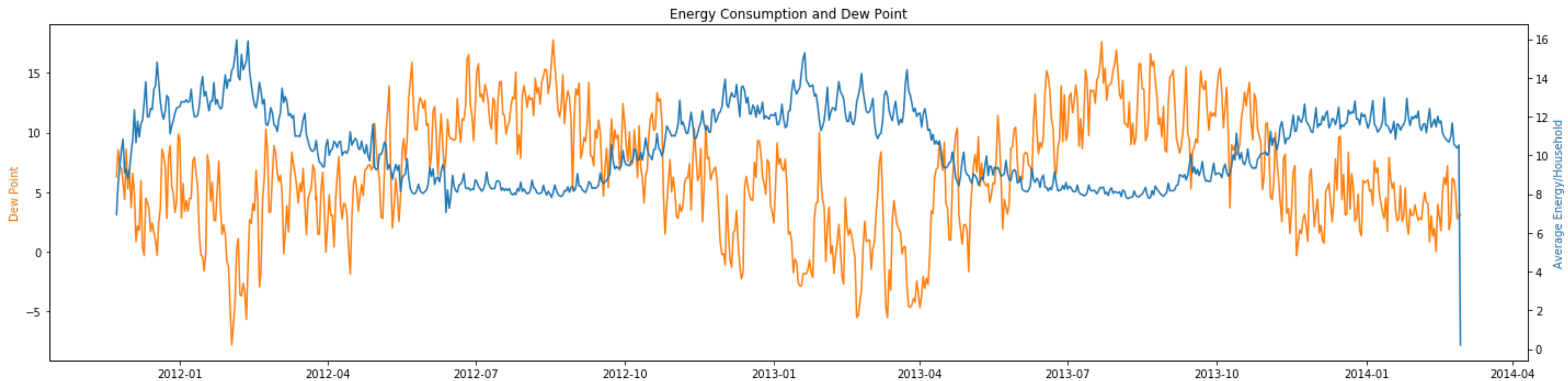


N.B. Y-axes are different to align data visualization

Observe relationship: UV index and Dew Point

- We observe that consumption has an inverse relationship with UV index and dew point





N.B. Y-axes are different to align data visualization

Observe linear relationship numerically using a correlation table

- Strong association if correlation >= 0.85
- Both variables increases if correlation > 0
- One variable increases and other decreases if correlation < 0

	avg_energy	temperature Max	dewPoint	cloudCover	windSpeed	pressure	visibility	humidity	uvIndex	moonPhase
avg_energy	1.000	-0.847	-0.756	0.242	0.150	-0.029	-0.246	0.361	-0.733	-0.032
temperatureMax	-0.847	1.000	0.865	-0.333	-0.154	0.119	0.259	-0.405	0.696	0.004
dewPoint	-0.756	0.865	1.000	-0.025	-0.092	-0.028	0.043	0.056	0.487	-0.008
cloudCover	0.242	-0.333	-0.025	1.000	0.170	-0.101	-0.330	0.480	-0.249	-0.062
windSpeed	0.150	-0.154	-0.092	0.170	1.000	-0.344	0.281	-0.042	-0.153	-0.023
pressure	-0.029	0.119	-0.028	-0.101	-0.344	1.000	-0.013	-0.251	0.101	0.038
visibility	-0.246	0.259	0.043	-0.330	0.281	-0.013	1.000	-0.578	0.240	0.063
humidity	0.361	-0.405	0.056	0.480	-0.042	-0.251	-0.578	1.000	-0.534	-0.014
uvIndex	-0.733	0.696	0.487	-0.249	-0.153	0.101	0.240	-0.534	1.000	0.013
moonPhase	-0.032	0.004	-0.008	-0.062	-0.023	0.038	0.063	-0.014	0.013	1.000

Now that we have a preliminary idea of the dataset, we can group similar observations by clustering and analyze them

- Why do Cluster Analyzing?
  - Finding groups in observations
  - Using some measure of similarity

- Types of clustering
  - Agglomerative
  - Divisive

- Common algorithms
  - K-means
  - DbScan
  - Hierarchical

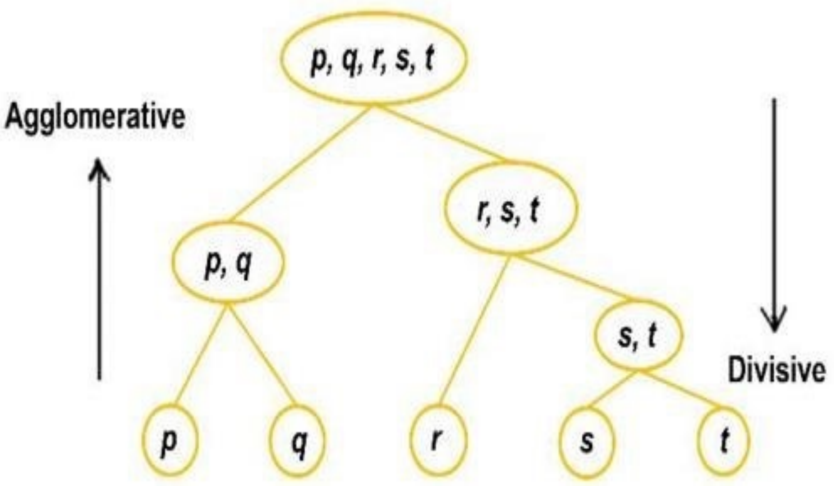
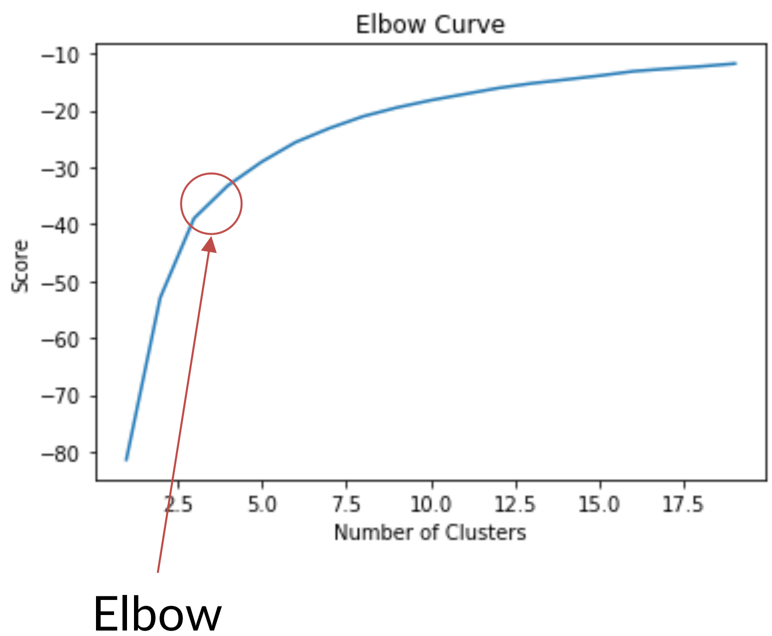
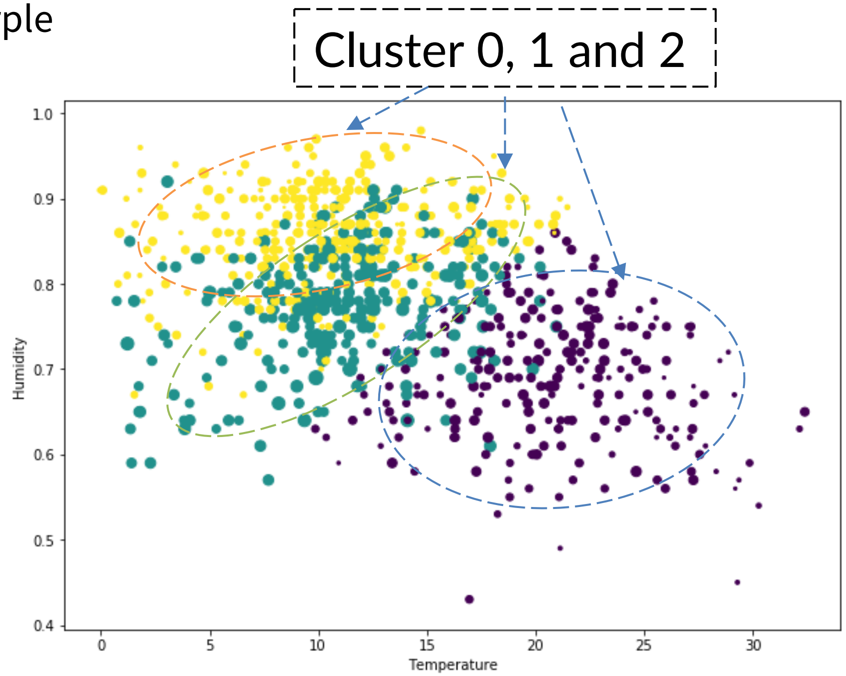


Figure Credit: [Giacoumidia et al. 2017]

Cluster weather data to group similar days

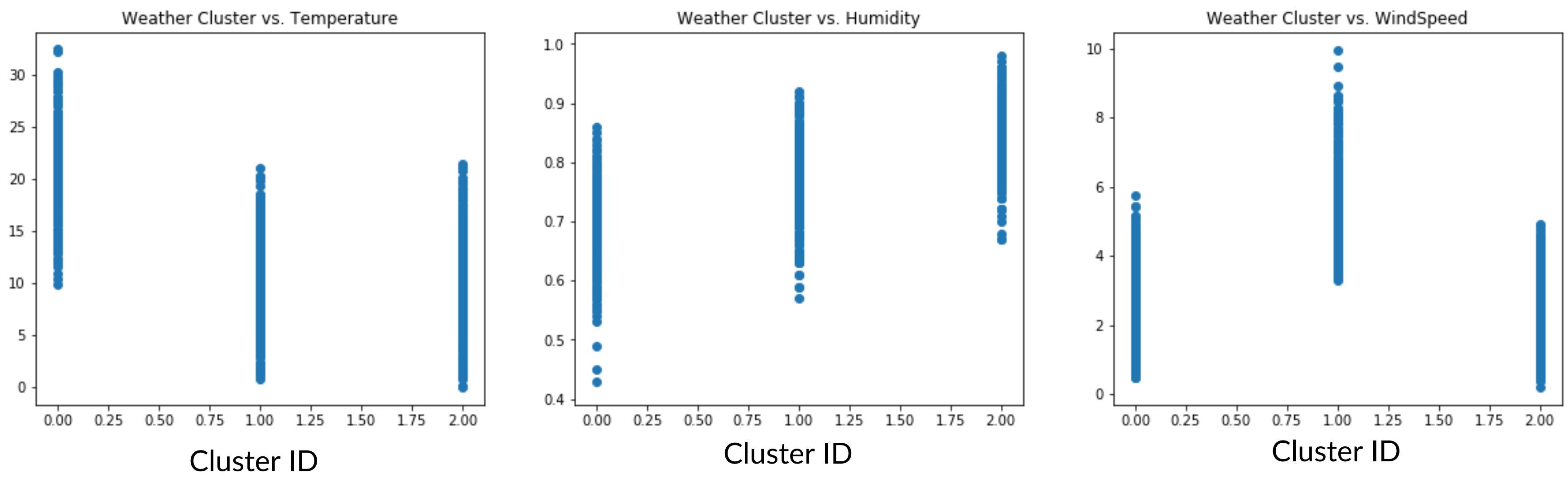
- Create clusters of weather data using K-means
  - Humidity
  - Temperature
  - Wind speed
- We use an elbow plot to determine number of clusters
- We get 3 clusters: colored yellow, green and purple



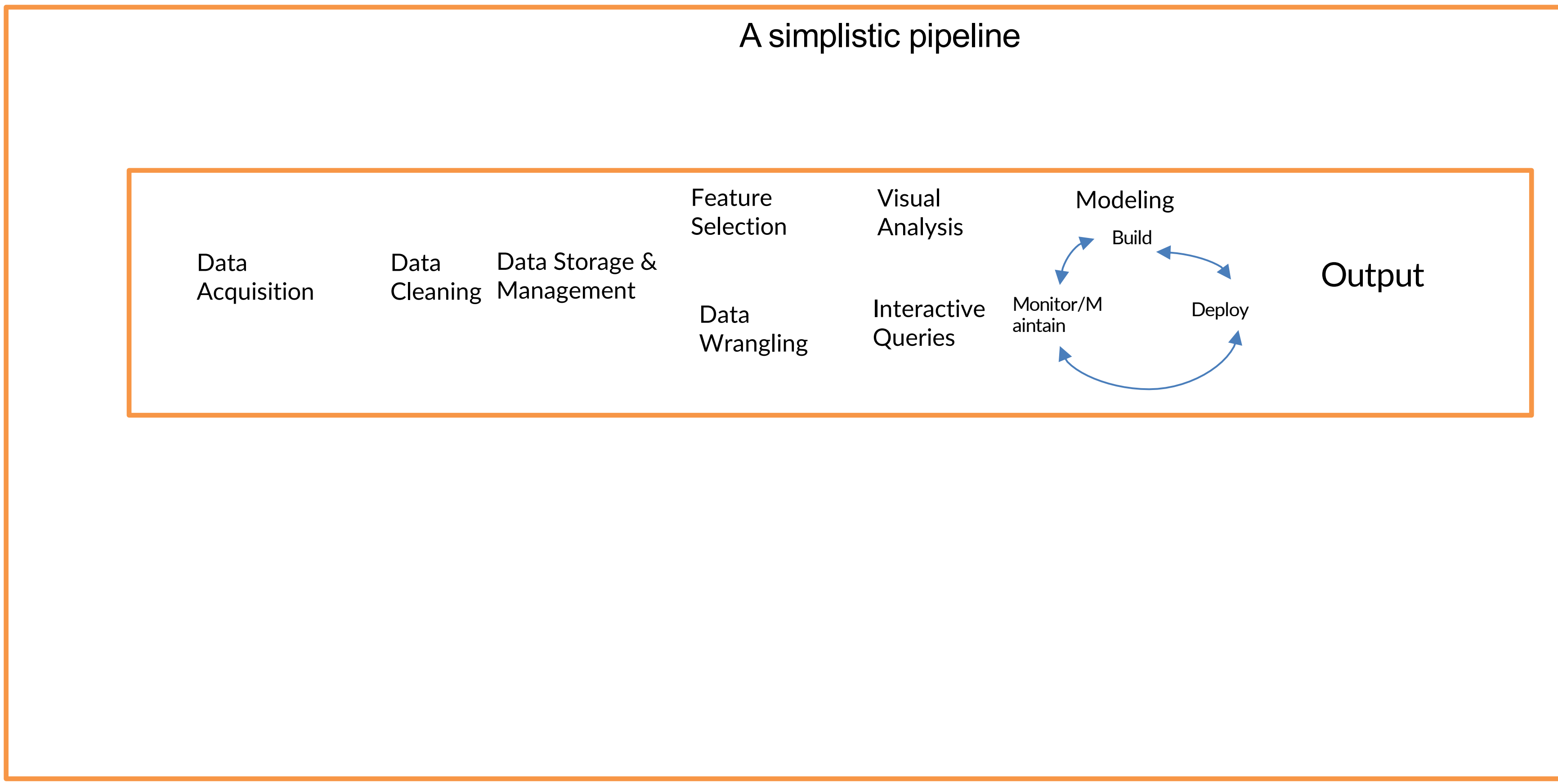


Relationship of weather variables in the clusters

- We observe humidity increases as we move from cluster 0 to 2
- But it is not the same in other variables
  - For temperature, only cluster 0 groups the hot days
  - Fore Wind speed, only cluster 1 groups the windy days
- So we might say
  - Cluster 0: contains the hot days with relative low humidity
  - Cluster 1: contains the windy days with relatively moderate humidity
  - Cluster 2: contains the high humidity days with relatively low temperature and wind

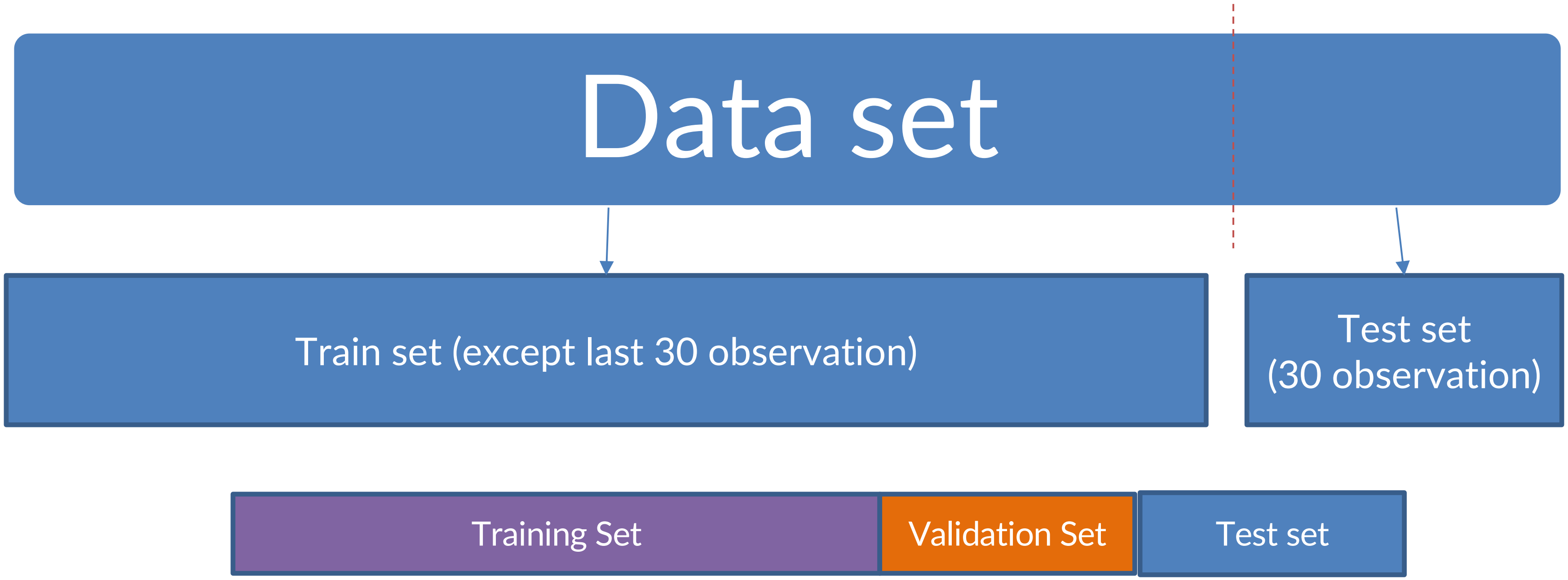


Simplified AI process



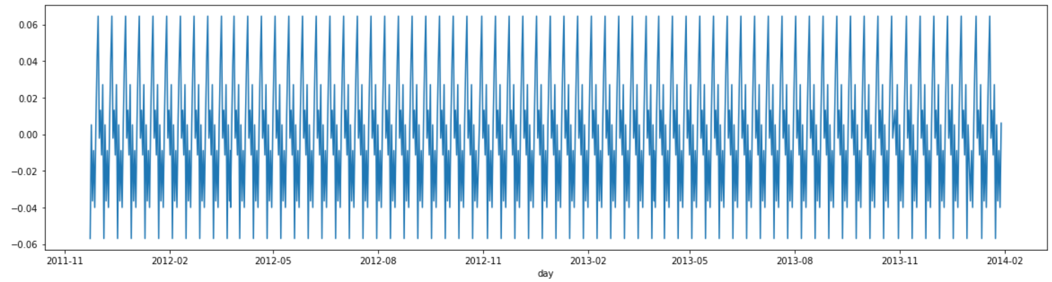
Preparing dataset for model training and test

- Split the data set into train and test set
- Train data for model train
- Test data for model evaluation

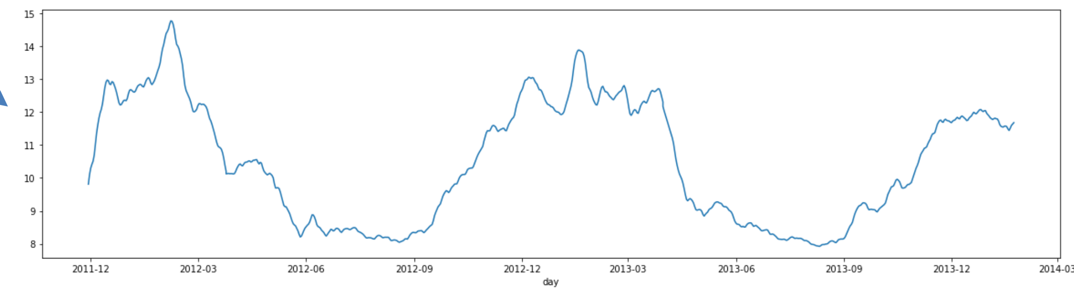


Statistical Models: Seasonal Decomposition

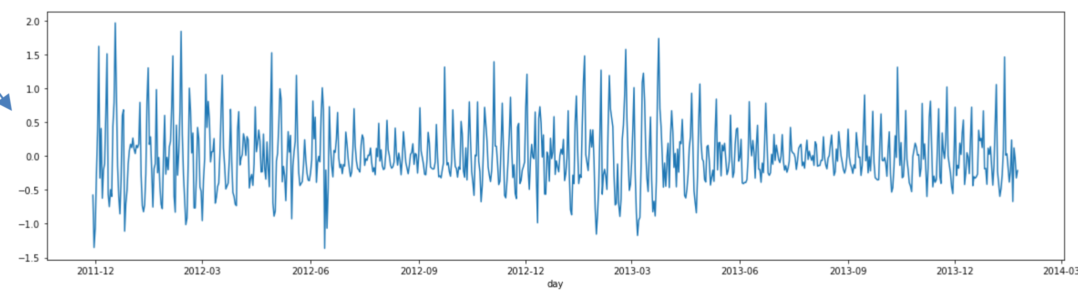
- Decompose time series into components
    - Seasonal
    - Trend
    - Noise



- Seasonal plot shows the recurring patterns in a week



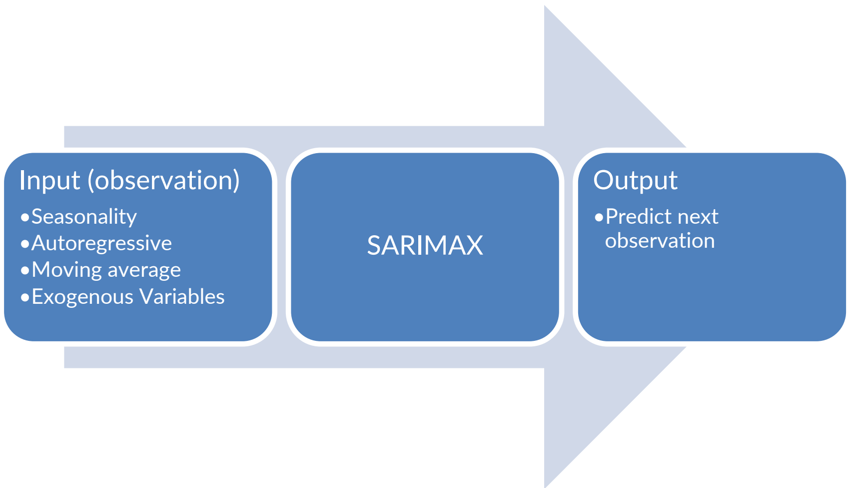
- Trends show changes over days
    - Higher consumption in winters for heating
    - Lower consumption on summer

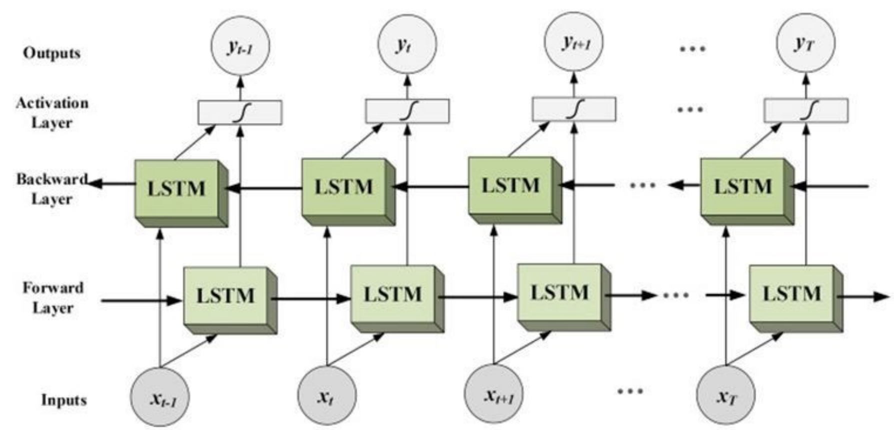


- Noise is the random variation in the data

Times series data can be predicted using SARIMAX and LSTM

- SARIMAX, A popular model for time series data with seasonality
- SARIMAX stands for Seasonal Auto-Regressive Integrated Moving Average with eXogenous factors
- SARIMAX is a shallow model
- LSTM is a recurring model architecture capable of storing long and short term information
- LSTM is a deep model





LSTM Image source: AnalyticsVidhya (link)

Over fitting

- **Overfitting** refers to a model that models the training data too well. **Overfitting** happens when a model learns the detail and noise in the training data to the extent that it negatively impacts the performance of the model on new data

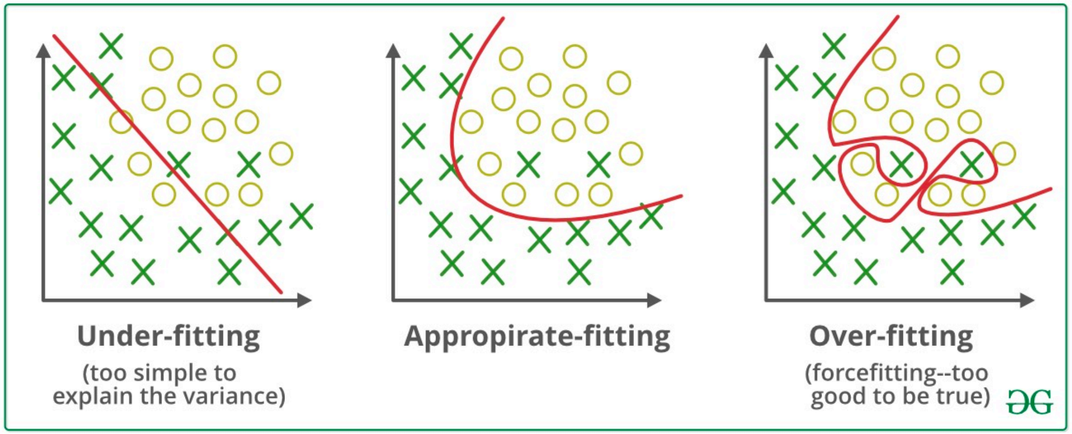
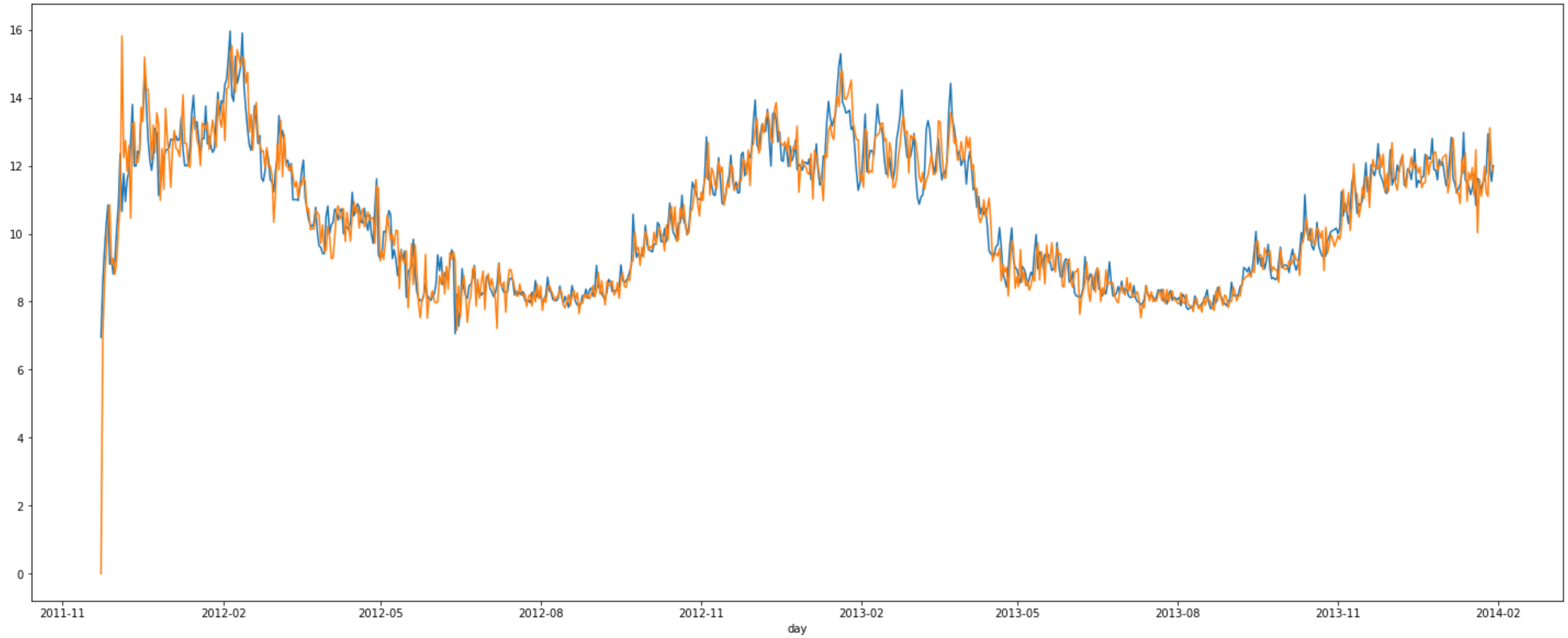


Image source : https://www.geeksforgeeks.org/underfitting-and-overfitting-in-machine-learning/

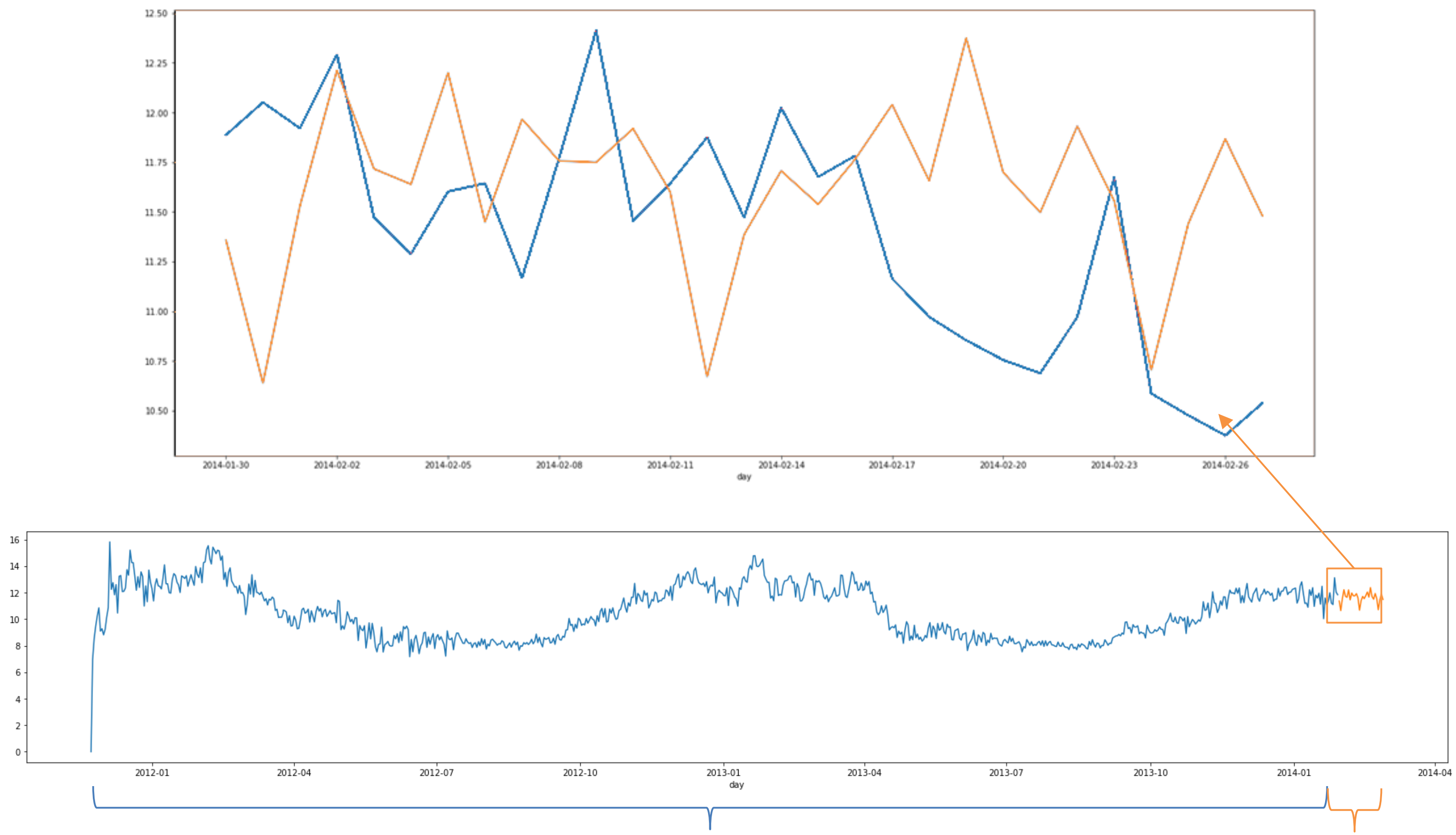
Training the model and checking fit

- In plot, blue is actual average consumption and orange is model fitted values



Testing prediction against actual values (test set)

Blue = actual, orange = prediction

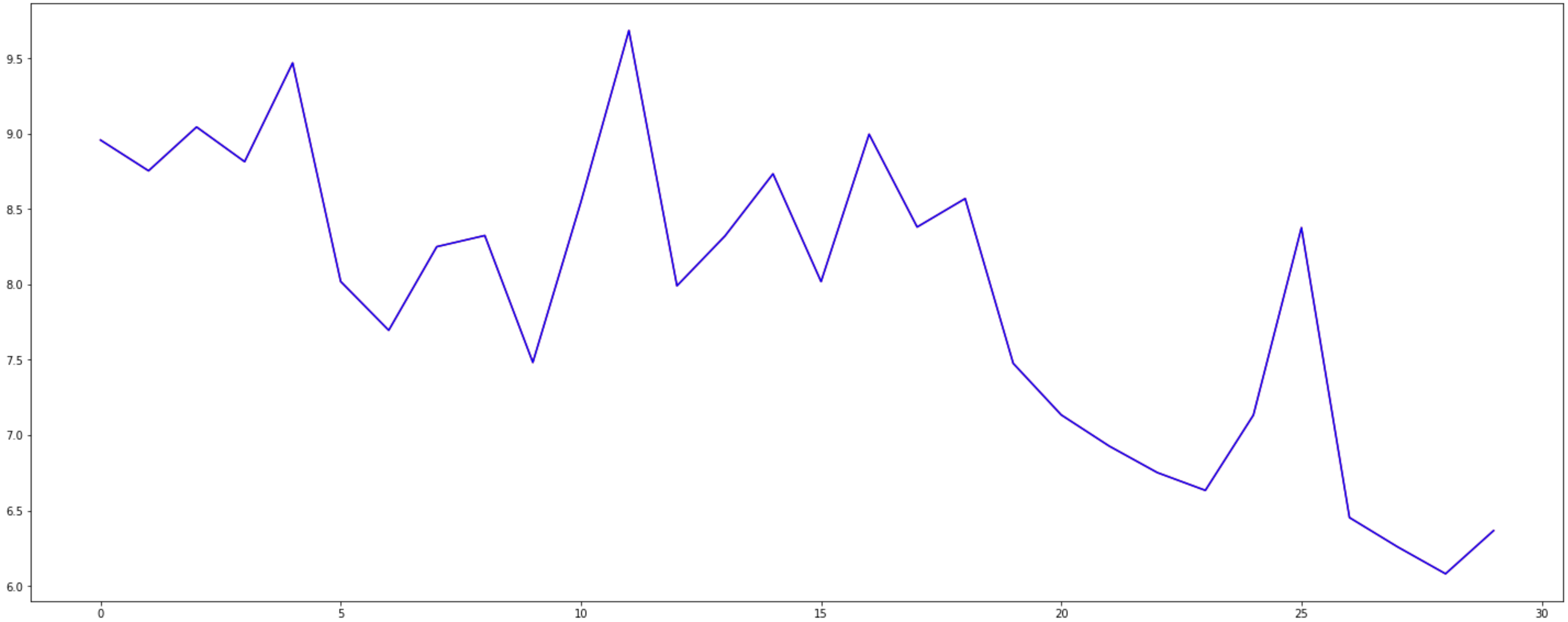


Train set

Test set

Deep model: Long short term memory (LSTM) network

- The model was trained using the same features for 50 epochs
- Plot shows actual as red and predicted as blue (overlapping lines)
- Test evaluation: Root Mean Square Error (RMSE): 0.000



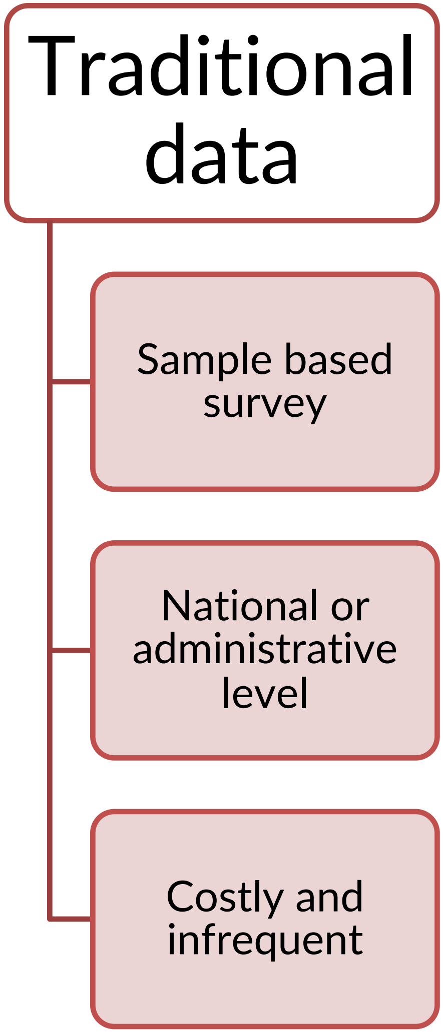
Smart use of AI and Data

Traditional data is costly, less frequent & reported on administrative level

- Traditional data sources are **sample based** and reported **on a national or administrative level**

- And is **costly**, for example, on average
  - Population census costs from 100s of million to US$12 billion [1]
  - Household survey can cost anywhere from US$ 460,000 to 1.7 million [2]

- That is why the **are not done frequently** and lacks sub-national resolution



- Sources:

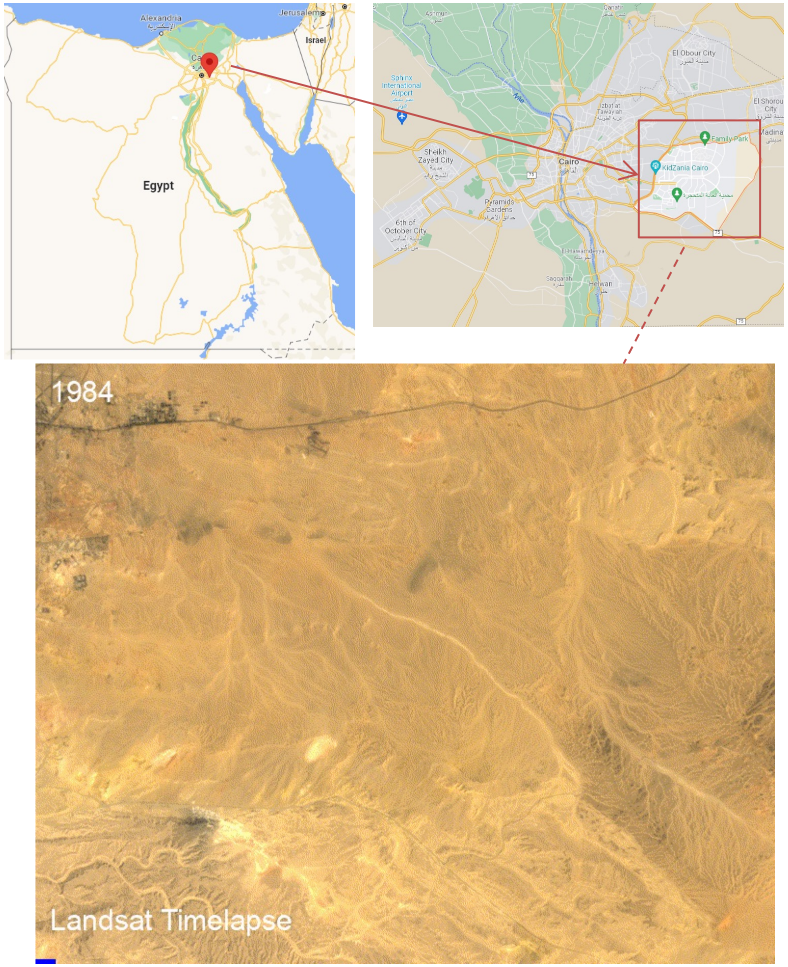
1. *Main Results of the UNECE-UNSD Survey on the 2010 Round of Population and Housing Censuses* (UNECE and UNSD, 2009).
2. Espey, J. et al. *Data for Development: A Needs Assessment for SDG Monitoring and Statistical Capacity Development* (UNSDSN, 2015).

Satellite can observe different bands which allows collecting various data at higher spatial resolutions

- Satellites can now fitted with equipment that can absorb different bands of light that can be used for various application

- For example,
  - **Landsat-8** offers images from 11 bands that absorbs wavelengths ranging from 0.433 to 12.5 micrometers
    - bands 2,3 and 4 are the visible blue, green and red lights
    - Source: [landsat 8, NASA.gov]

  - Joint Polar-orbiting Satellite System (**JPSS**), fitted with the Visible and Infrared Imaging Suite (VIIRS) Day Night Band (DNB) provides astounding improvement on low light imaging ([Earth Observation Group])



Landsat (visual color bands) for land development

GIF shows the yearly time-lapse of Cairo New City (Source: Q. Wu, streamlit.gishub.org)

LandSat8 Bands and their combination for uses

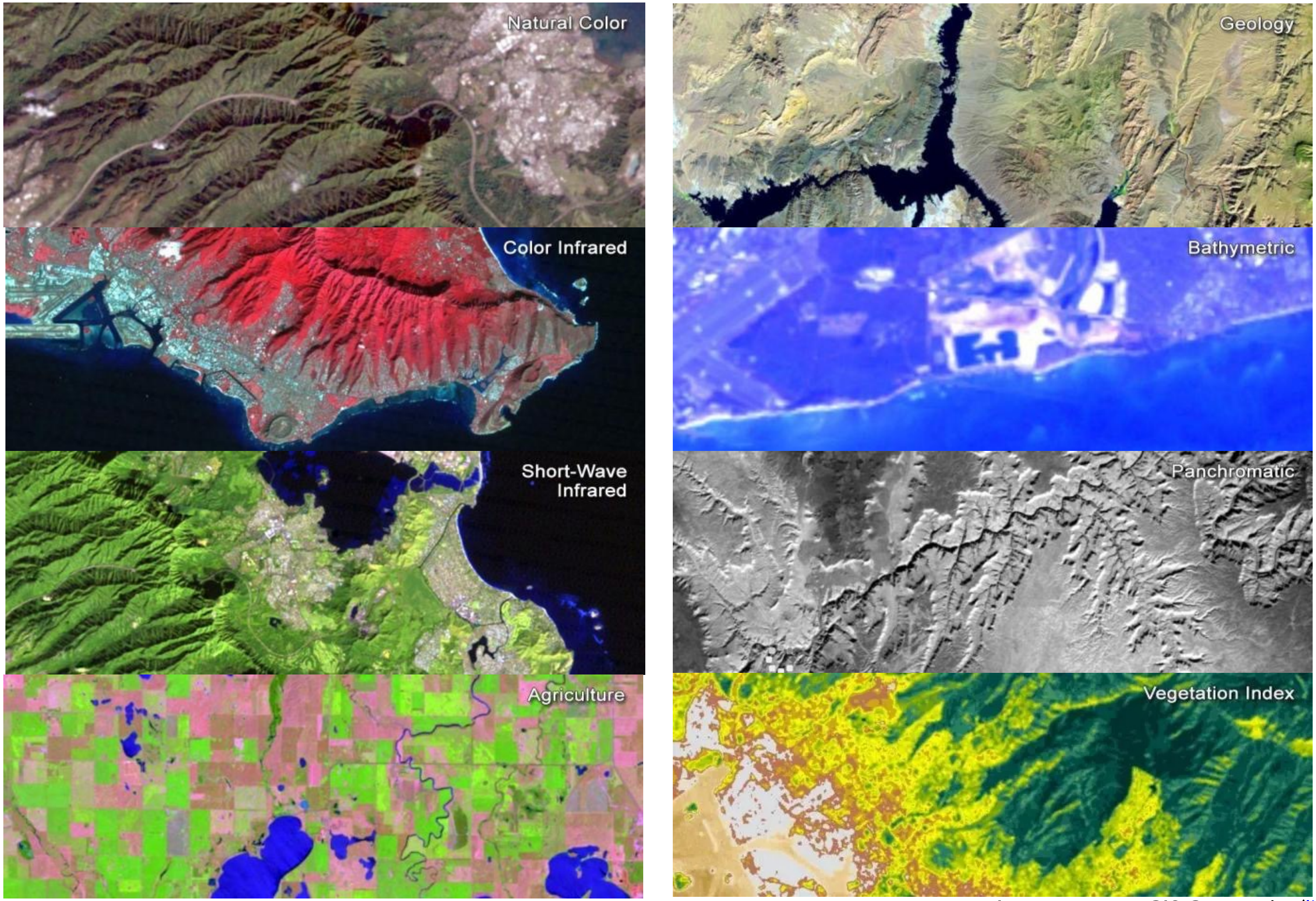


Image sources: GIS Geography (link)

JPSS: Monitoring Nighttime lights for studying socio-economic factors

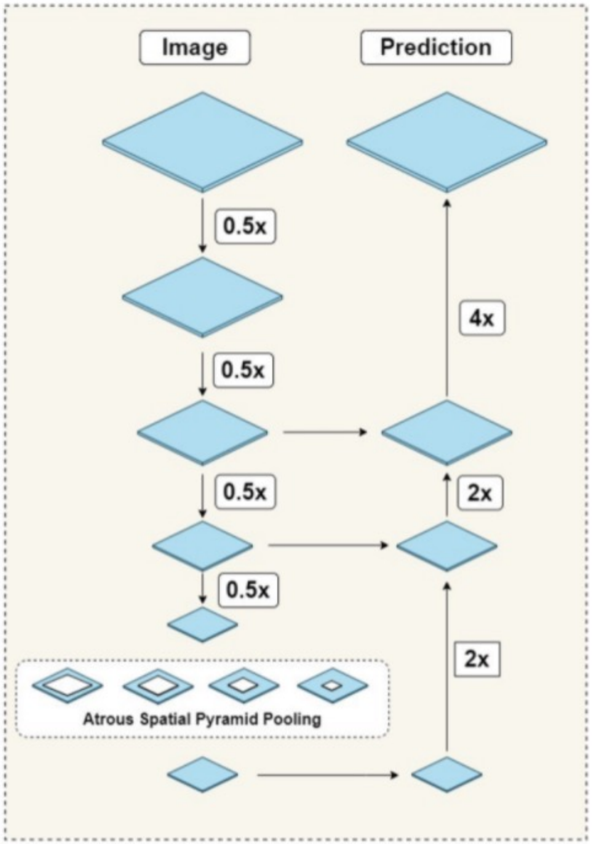
- Electricity consumption is highly correlated with
  - Wealth
  - Health outcome
  - Education
- Continuous data
  - Every month (NOAA)
  - Every day (NASA)
- Disaggregated data
  - Sufficient number of samples at sub-national levels

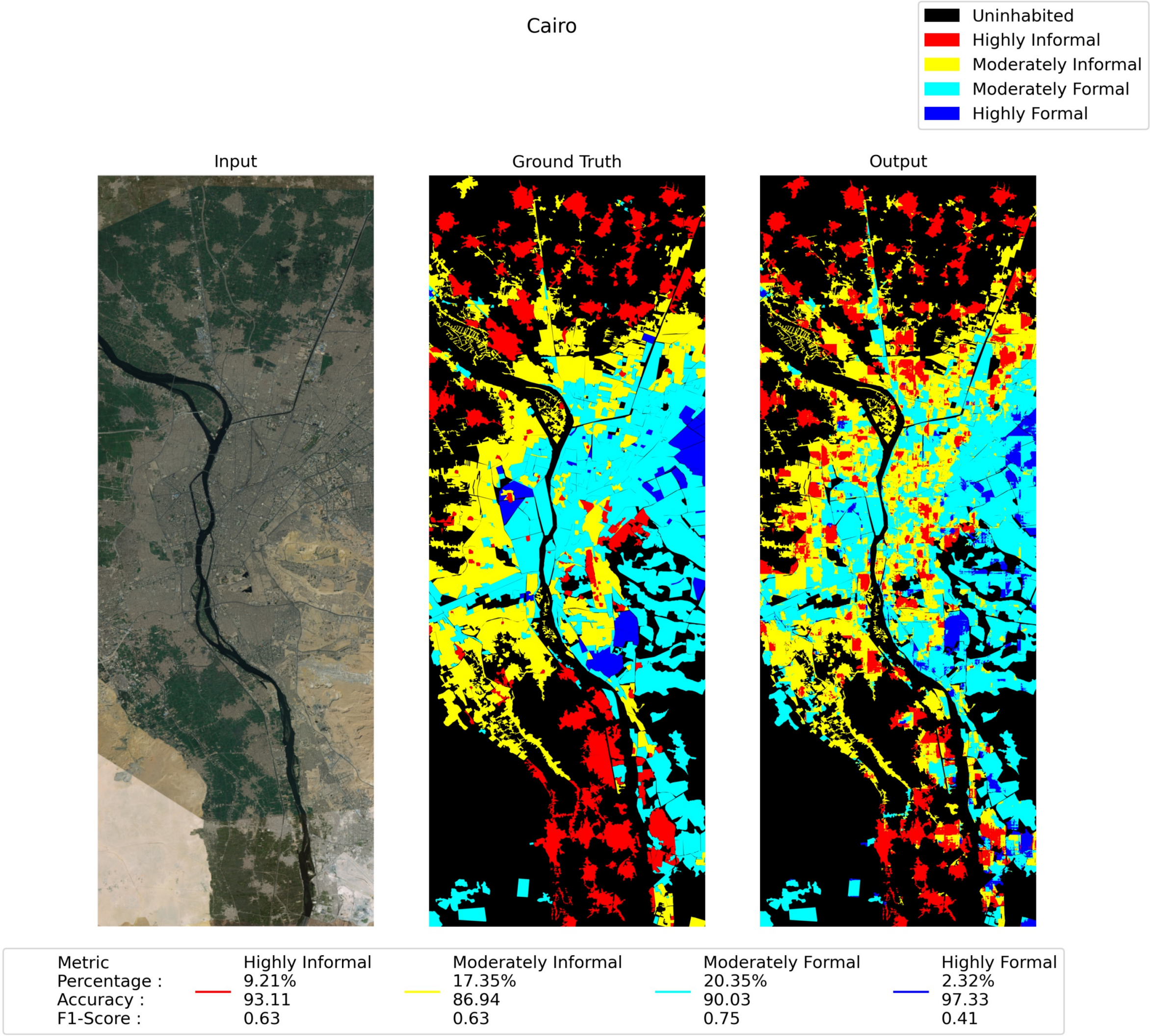
- Can be use for
  - “Immediate” change in intensity of economic activity as outcome of interventions/projects
  - Across group variations in outcome of interventions
  - Groups based on wealth index, income quantile based on DHS, HIES
  - Measuring inequality
  - Per capita income or per capita consumption from HIES
  - Monitoring impact of natural disaster or external shock

- Source: (Wahed et al. 2020)
- Dataset available from Earth Observation Group (link)

Source: Black Marble (VIIRS, SNPP), image optimized for print

Using deep models to categorize land cover and use

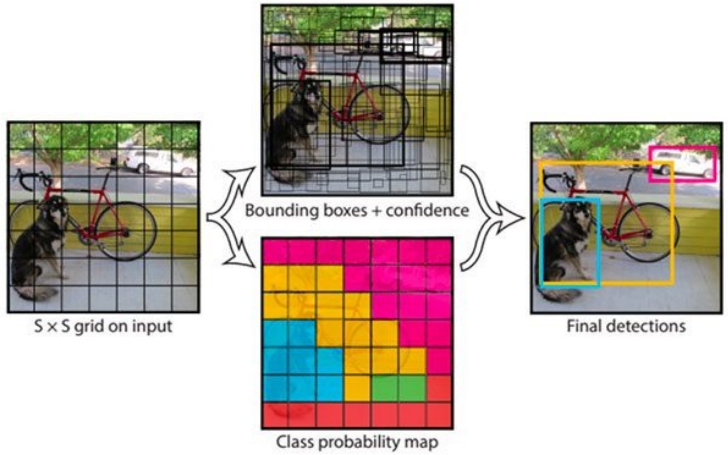




Source: dndlab.org, Paper link (link)

Car number plate extraction from road cameras using YOLO model

- Traffic camera is a commonplace for many cities and is a given for a smart city
- Automated car plate extraction is used to create transport analytics
- This is done by Yolo model which stands for “You Only Look Once”

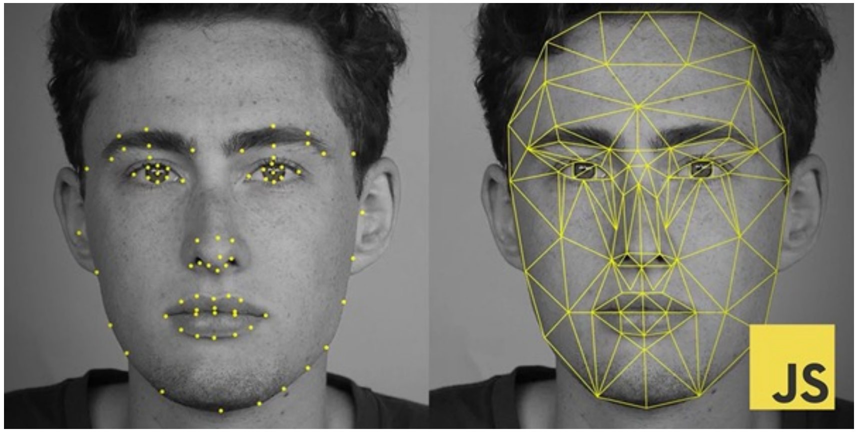


Yolo object detection steps (Redmond et al. 2016)

Yolo for car plate detection

Face recognition by extracting and comparing faceprint from the nodal points

- Face recognition is a common feature in making computer vision packages
- It use the nodal points from faces
  - width of your nose;
  - depth of eye sockets;
  - shape of the cheekbones;
  - distance between eyes;
  - And others
- To create a “faceprint” that can be stored and used later for comparing
- Public safety benefits
  - monitoring events
  - finding missing people
  - Helping investigations



Face recognition using nodal points (Source: skybits)

Face recognition allows tracking citizens in the streets (Source: Arteco)

Recognition is being adopted in large scale which stores and uses face data to work and improve accuracy. But did they take public consent?

- There are many businesses who offer recognition platforms to police, security and other departments who use them to provide
  - better public safety by police
  - automate services such as auto-check-in and payments in subways
- Doing so, face data of countless citizens are being stored in servers which the citizens never consented to
- In addition, these platform providers also want to collect data "from the wild" to improve their services
  - Which means they might also retain facial data
- Source: Roussi, 2020, Nature (link)

Subway station in Zhengzhou, China
Photo Credit: Gilles Sabrié/NYT/Redux/eyevine

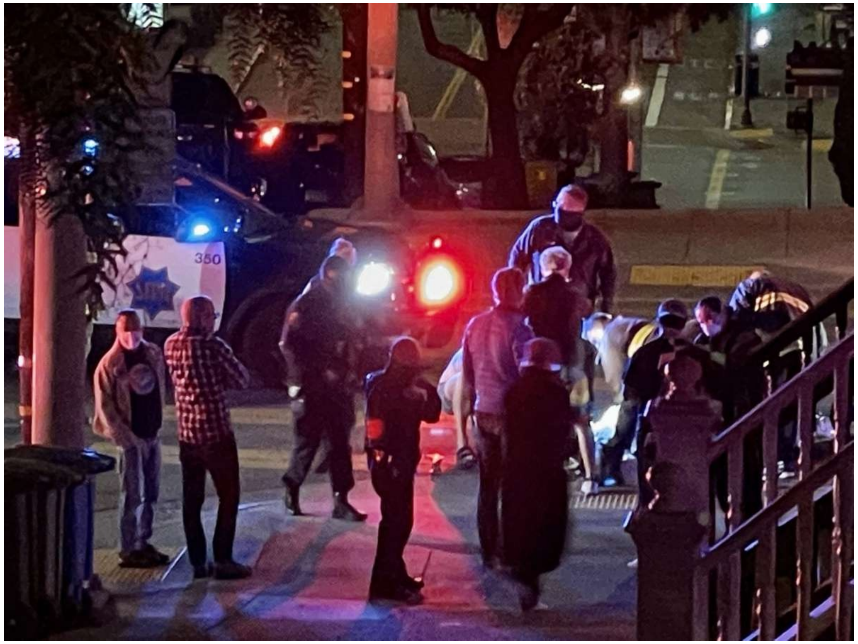
But they can also be used to keep tabs on citizens

- according to CNET (author, date, link) is using face recognition to track citizen and their activity
- Add or reduce citizen's "social credit" based on certain events
- Incentivize high credit with benefits while punishing bad credit

Face and car recognition in a street according to CNET (source: CNET)

Accurate location information is need to coordinate police resources after citizens report a crime but gunshot sources can be hard to pinpoint

- Often times citizen may not be able to call 911 as fast as needed
    - Possible reason could be a state of shock due to startling sound of gunfire
- Location is hard to pinpoint acoustically by humans in a city
- Connecting a call and providing an accurate description can take a few minutes
- And call center a minute or two to reach nearby police who can respond
- Unfortunately all these takes up precious time which may be used to help the victim

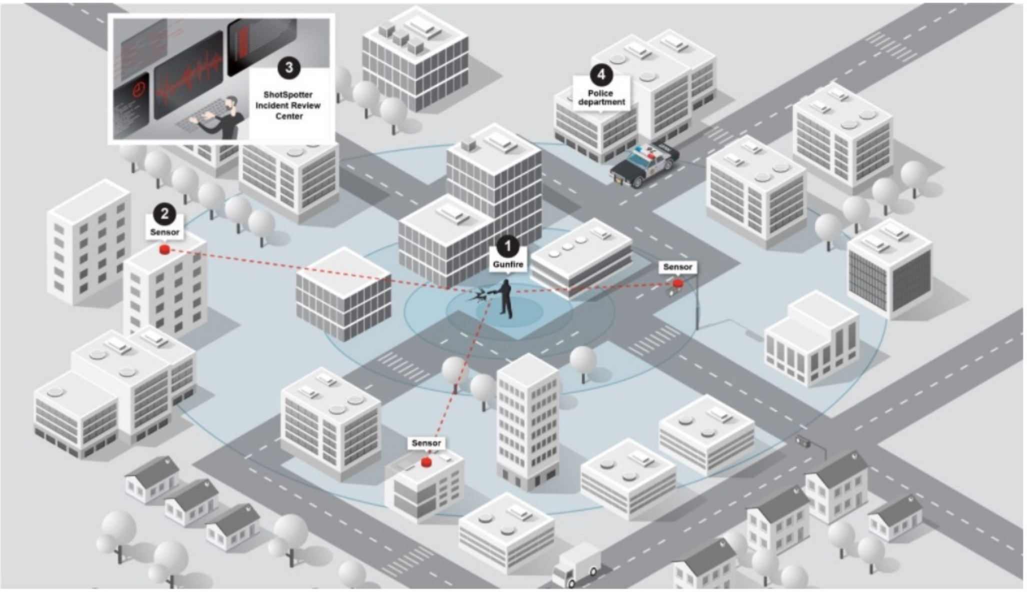


Shooting of a man in San Francisco's Noe Valley during robbing incident. Photo credit: Noe Valley resident, (from SFGate)

Using multiple acoustic sensor to triangulate source and use of AI to detect shot and notify police takes less than 60 seconds

- ShotSpotter is a company that focuses on precision policing technology solutions
- They have development a platform capable of detecting gunshot accurately as well as the location of the gunfire using sound and AI
- The process also can notify police and takes less than 60 seconds
- They are being used by San Francisco Police Department to provide faster response times and decrease violent crimes
- Source: Cassidy, 2020, sfchronicle.com ([link](#))

How Gunshot Detection Works



1 Gun is fired

When a gun is fired, the sound of a muzzle blast radiates outward.

2 Gunshot is Detected and Located

Acoustic sensors are triggered by the impulsive sound. THe sound is classified as a gunshot using artificial intelligence and triangulation determines the precise location.

3 Gunshot is Reviewed

The data is relayed to the Shotspotter Incident Review Center where analysts quickly audit the data and publish confirmed gunsots to police.

4 Police Respond

Alerts are sent to dispatch centers and patrol officers' smartphones and MDTs for immediate response. The entire process takes less than 60 seconds.

Source: [ShotSpotter](#)

Data from ride-hailing services are used for transportation planning, while constructive, it is a privacy concern since it tracks every citizens

- Ride sharing Company takes note of vehicles, where they went and how they got there, then sends that information to local officials.

- The idea is to allow cities to keep tabs on how well companies are adhering to regulations
    - commitments to serve low-income neighborhoods
    - or to get info on where to plan future bike lanes.

- But all these data also contains where citizens are going and infringes on personal privacy rights

Hailing a ride using the mobile application.
Photo credit: Rajib Dhar/Dhaka Tribune

We should aspire to reach a collective agreement balancing the beneficial use of data in new technology while also respecting personal privacy

- “Technology may have made it easier to measure urban life, but it doesn’t mean we’ve reached a collective agreement about what aspects of urban life should be measured, or by whom” says Molly Turner, a startup policy expert and advisor to Spin (a bike sharing startup)

- New technology and climate concerns are likely to cause fundamental changes in the ways people get around cities, and planners are going to need “unfettered data” about those changes to deal with safety, equity, quality of life, and other issues, says Connie Llanos, LA Department of Transport’s chief of staff and assistant general manager.

Thank you for the participation

Email: zaber@du.ac.bd