

Familiarity with Python

- 50 % hv some python knowledge
- Use excel + python + whiteboarding

- Data collection

- Wide range of variables
- Viewer
- Viewer behaviors
- Show content
- View time
- Age.... Demographics
- Time spent
- Switching time
- Fwd/prev
- Temporal (date, time ...)
- Viewer rating, responses ..
- Industry data
- Historical behavior

- Data quality

- Completeness
 - Biz rule/requirement driven
 - Enough values
 - Drug names
 - ◆ Are we covering all the drug names
 - Dimensions of values (cardinality)
 - Attributes
 - Properties
 - Examples
 - ◆ Drug types (d1, d2, ...d99)
 - ◆ Vehicle type
 - ◇ (sedan, SUV...convertibles)
 - Category
- Data truncation

○ Cleaning

- Duplicates
- Missing data
- Formatting
- Nulls
- Date formats
- Type mismatch

○ Noise/outliers

- Invalid values as noise
 - Unusual deviation
 - Ref to MIN, MAX, RANGE, STD DEV, VAR

....

- Anomalies (application)**
 - Banking - transaction (reference to normal patterns)
 - Medical MRIs (reference to normal MRIs)
 - images
 - Require some ML models

- Accuracy of the data

- Exploratory data analysis (EDA)

- Viz
- Describe the data
 - Var, std dev, min, max
- Distribution of data
 - Symmetry
 - Clustering/groups
 - Dispersion
- Inferential statistics**
 - Age vs viewing time
 - Genre vs viewing time

CC

↓
X₁
y
N
—

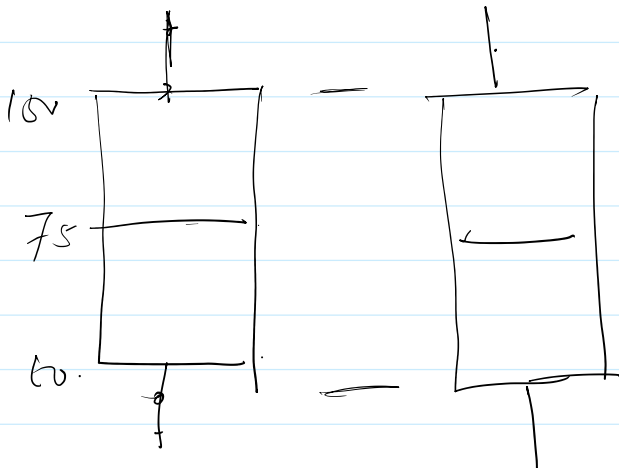
↓
X₂
4WD
RWD
FWD
—
Categorie

X₃
:
:

Cars

↑
testing
time
90 s
110 s
137 s
:
:
:

4WD



desirable?

4WD 2WD Inferential statistics (categorical)

Modeling and predictions

- Model
 - o Variety of approaches
 - Statistical
 - Machine / deep learning
 - Simulation
 - Visual rep
 - o Select the optimum model(s)
- Deploy and predict
 - o Cloud , in premise

Multiplicity of causes

- Implications
 - o Not a single one-to-one relations
 - o Identifying the contributing factors
 - How to handle
- Solutions
 - o Regression analysis
 - o Multivariate analysis
- Examples of projects where this is not that much of problem
 - o Machine problems
 - Assignment
 - Scheduling
 - Maintenance
 - o Medical projects

Qualitative data

- Non-numerical
- Examples
 - o Categories
 - o Feedback (text)
- Analysis
 - o Viz
 - o Freq based measures
 - Counts, probs ..
 - o Hypothesis testing

Quantitative data

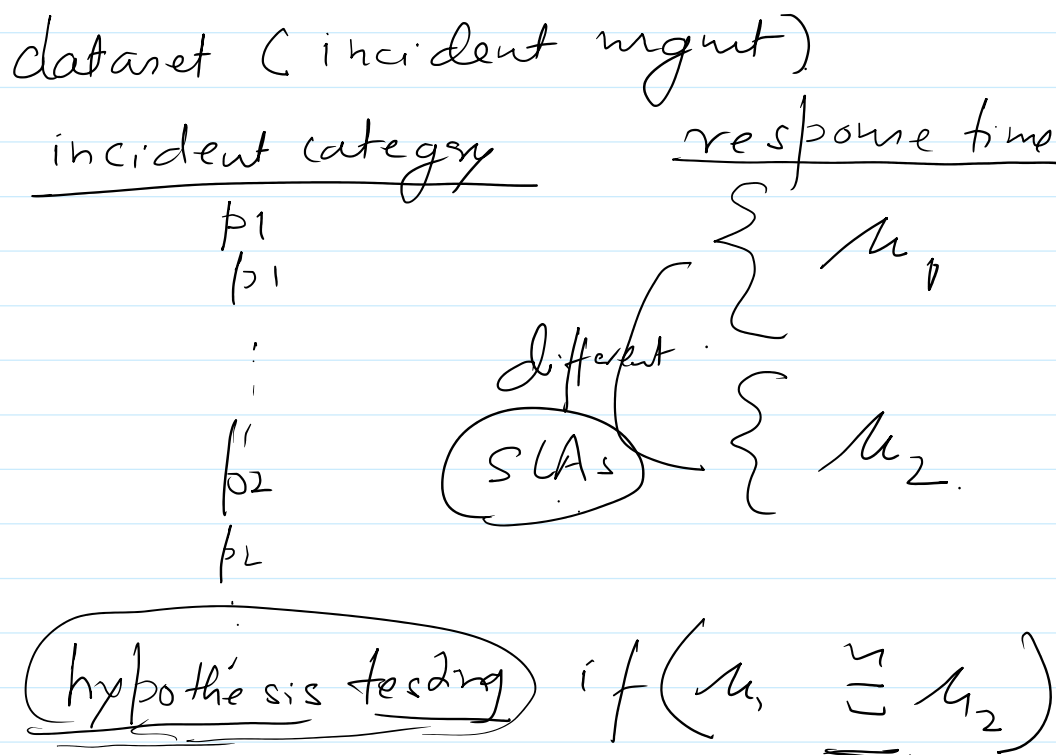
- Numerical in nature
- Measurable
- Mathematical operations
- Example
 - o Training time, age....
- Analysis
 - o Summarize
 - o Relations/COV
 - o Infer from the relations
 - Regression analysis
 - Hypothesis testing

Why **generalization** is difficult?

- Randomness of trg data
- Unseen data
- Not possible to get 100% view of the data
- **Complexity**
 - o Columns or values in the columns
- **Biased** data

Predictions

- ML/DL type of model
- Statistical approach (inferential methods)



incidence

type appln

Custom

(?)

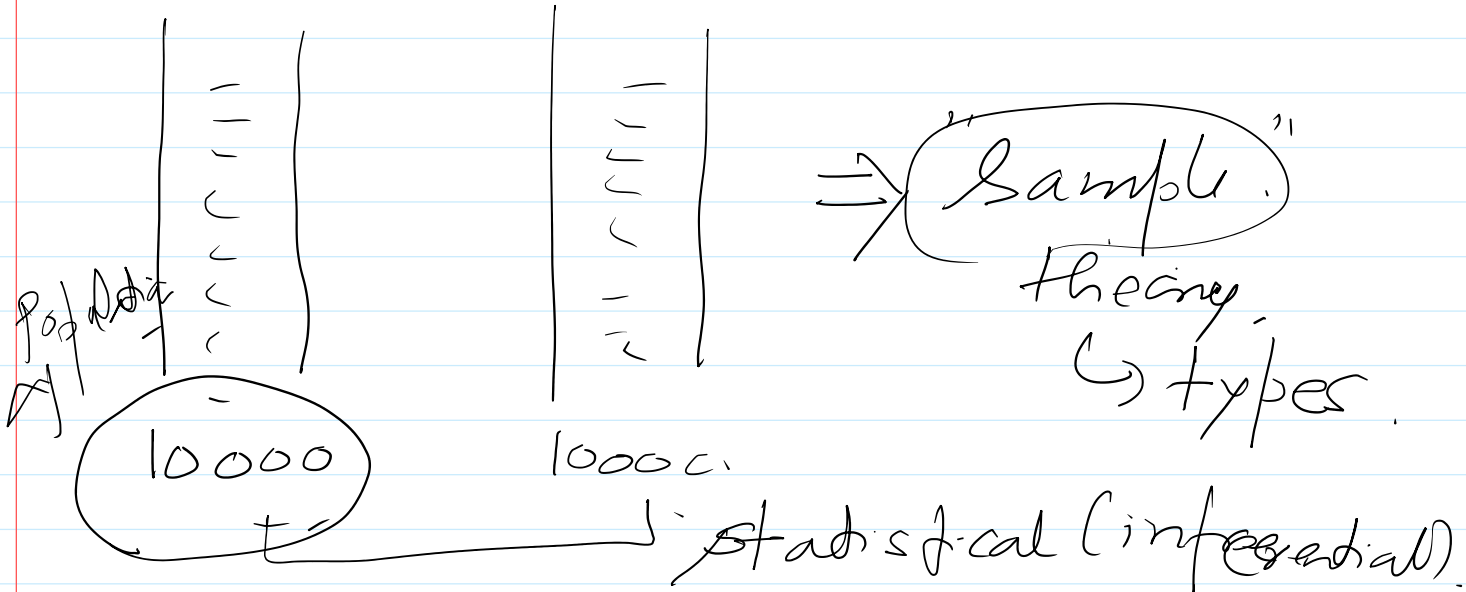
Time to resolve



(ML/DL)

Compare two marketing schemes.

$e_1 \approx e_2$



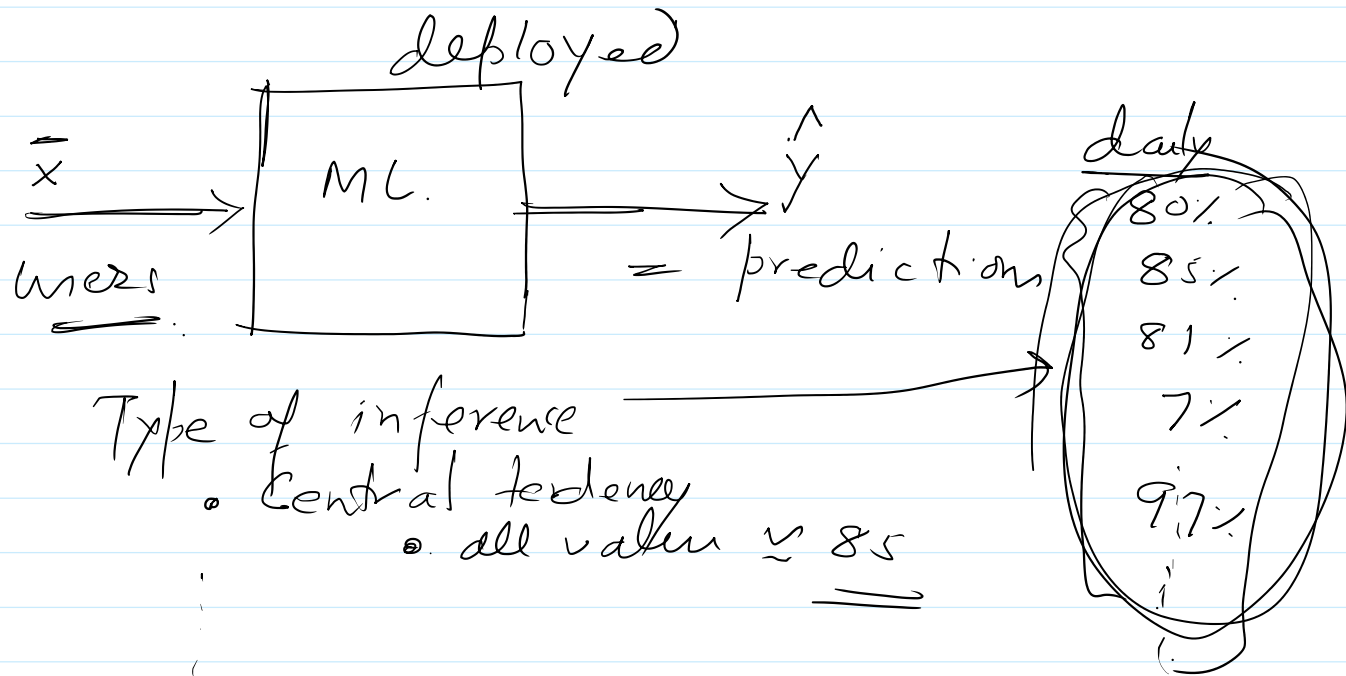
Descriptive analytics

Predictive analytics (ML, DL ... ADL... RLCV, text, LLM

Prescriptive analytics

- Projection
- Recommendation
- Suggestions
- Example
 - o 5 engg to work on tickets
 - Different skills
 - Cost rates
 - Availability calendars
 - o Weekly forecast of 1000 tickets to handle
 - o How u recommend a roster?
 - To predict if we can handle all the 1000 tickets

- How u recommend a roster?
 - To predict if we can handle all the 1000 tickets



Calculus/understanding in your 44 weeks

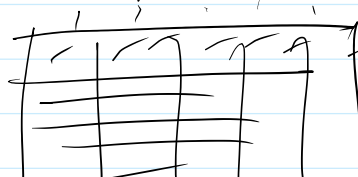
- class/grade 11 (maths).
 - differential (first order)
 - grade 9th/8th
 - distances (euc - /variance)
 - sets/...
 - Integration calculus (basic)
- sklearn
pytorch
TF/Kera
- |||||

Understanding

- Varianles
- COV
- distribution.
- pres. . . .

\Rightarrow matrices?

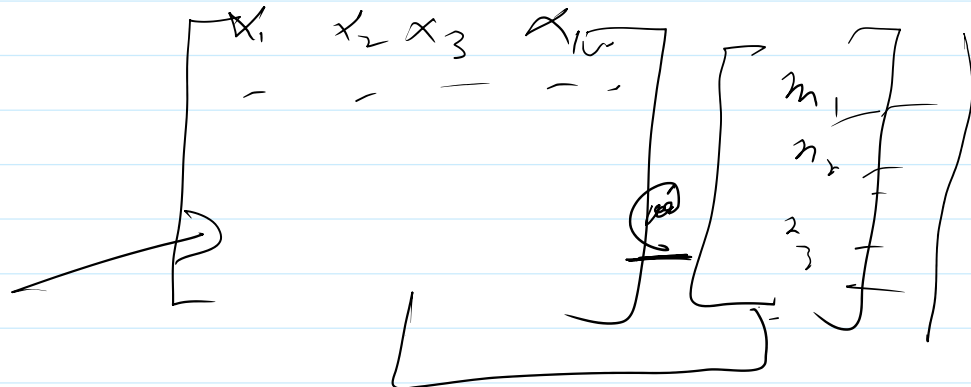
example:





$$y = mx + c$$

$$= m_1 x_1 + m_2 x_2 + \dots + m_{100} x_{100}$$

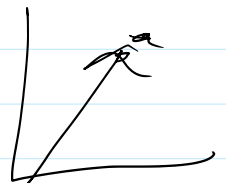


AI (linear transfer)

Why combining cols \Rightarrow

- alters data
- explainability $\uparrow \downarrow$
- "interaction"

age ratio
37 5



Data table

- Rows and cols
- Rows
 - Record, instances, **point**, case, entity, Objects, **document**, transaction,
 - Feature vector [37, 7, 5, ...]
- Cols
 - Features, fields, attrb, characteristics, dimensions
- Columns
 - Dependent and independent columns
 - Dataset
 - Temp-1 day before
 - Temp-2 days before
 - Avg temp /week
 - hum

- Wind pressure
- ..
- ..
- ACTUAL TEMP
- Independent columns

Sample dataset (temp forecasts)

year	month	day	week	Temp 2 days before	Temp 1 day before	Average temp	Actual temp on that day	Temp forecast by noaa	Temp Forecast by acc	Temp forecast by friend
2016	1	1	Fri	45	45	45.6	45	43	50	29
2016	1	2	Sat	44	45	45.7	44	41	50	61
2016	1	3	Sun	45	44	45.8	41	43	46	56
2016	1	4	Mon	44	41	45.9	40	44	48	53
2016	1	5	Tues	41	40	46	44	46	46	41
2016	1	6	Wed	40	44	46.1	51	43	49	40
2016	1	7	Thurs	44	51	46.2	45	45	49	38
2016	1	8	Fri	51	45	46.3	48	43	47	34
2016	1	9	Sat	45	48	46.4	50	46	50	47
...
2016	1	19	Tues	50	54	47.6	48	47	49	53
2016	1	20	Wed	54	48	47.7	52	44	52	61
2016	1	21	Thurs	48	52	47.8	52	43	51	57

Independent variables

Independent variables

Dependent variable