



Hyperlocal Weather Nowcast

Singapore

General Assembly DSIF4 Capstone Project
Lee Zhen Ming

TABLE OF CONTENTS

01

Problem
Statement

02

Data
Collection

03

EDA

04

Feature
Engineering

05

Scoring &
Baseline

06

Data Modelling



01

Problem Statement

Why do we need weather forecasting?

Being situated near the equator, Singapore has a tropical climate, with abundant rainfall, high and uniform temperatures, and high humidity all year round.



Given the high rainfall all-year round, this poses a risk to workers working in outdoor conditions.



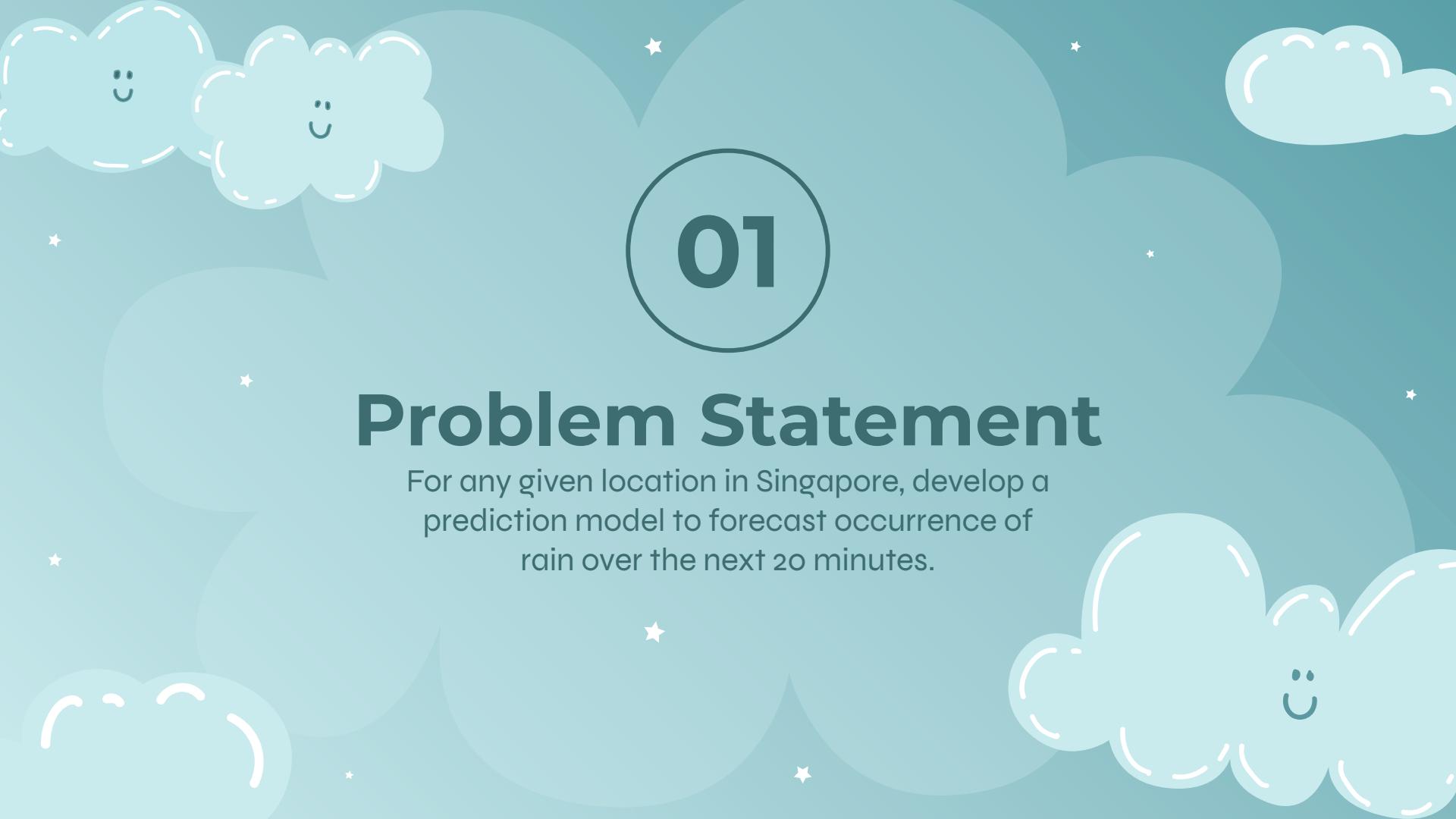
To mitigate this risk, project managers tend to postpone their work, leading to project delays and productivity losses



Image ID: AHF25A
www.alamy.com

Having rain prediction capabilities would enable project and on-site managers to plan better and manage workflow to minimize risk.





01

Problem Statement

For any given location in Singapore, develop a prediction model to forecast occurrence of rain over the next 20 minutes.



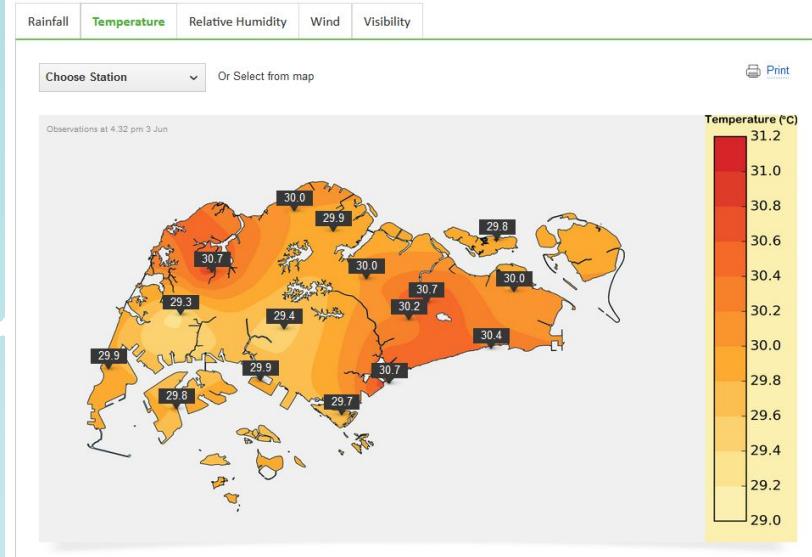
02

Data Collection

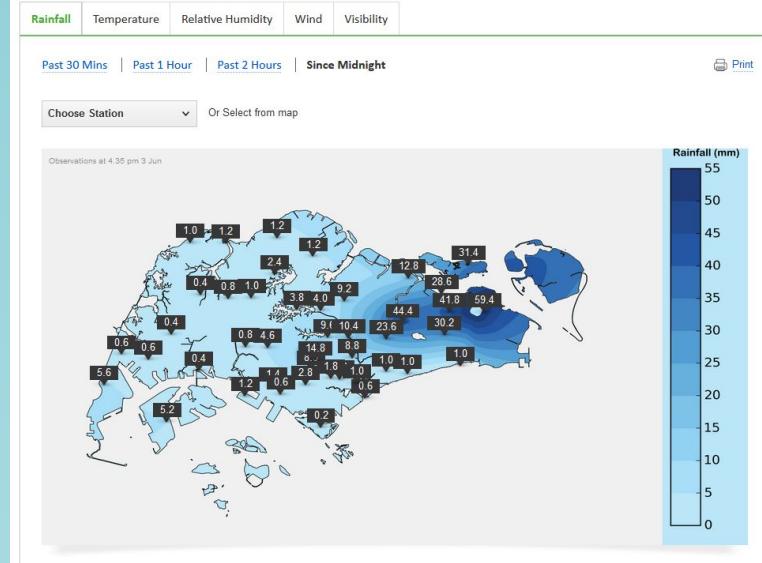
No data, no talk.

Data Sourced from data.gov.sg

Current Observations



Current Observations



Data Sourced from data.gov.sg

The screenshot shows a dataset titled "Realtime Weather Readings across Singapore". The left sidebar lists five datasets: "Air Temperature across Singapore", "Rainfall across Singapore", "Relative Humidity across Singapore", "Wind Direction across Singapore", and "Wind Speed across Singapore". The main content area displays the "Air Temperature across Singapore" dataset. It includes a "Views" section with a gear icon and a map icon, and an "API View" button. Below this is a detailed API endpoint for air temperature:

GET https://api.data.gov.sg/v1/environment/air-temperature Get air temperature readings across Singapore

Has per-minute readings from NEA
Use the `date_time` parameter to retrieve the latest available data at that moment in time
Use the `date` parameter to retrieve all of the readings for that day.

Parameters

Name	Description
<code>date_time</code>	YYYY-MM-DD[T]HH:mm:ss (SGT) string (query)
<code>date</code>	YYYY-MM-DD string (query)

Responses

Minute-by-minute air temperature readings at weather-station level.

Source: Data.gov.sg

```
},
{
  "timestamp": "2022-06-03T00:03:00+08:00",
  "readings": [
    {
      "station_id": "S109",
      "value": 28.3
    },
    {
      "station_id": "S50",
      "value": 27.8
    },
    {
      "station_id": "S107",
      "value": 29.3
    },
    {
      "station_id": "S43",
      "value": 28.3
    },
    {
      "station_id": "S108",
      "value": 28.5
    },
    {
      "station_id": "S44",
      "value": 27.7
    },
    {
      "station_id": "S121",
      "value": 27.7
    }
  ]
}
```

Well structured data

Weather Data Collected



Temperature

Sensible heat. Measured in degC.
1min per observation



Wind speed / direction

Surface wind speed and direction.
Measured in km/h and azimuth.
1min per observation



Relative Humidity

Amount of moisture in the ambient air.
1min per observation



Rainfall

Amount of rain collected over an area. Measured in mm.
5min per observation

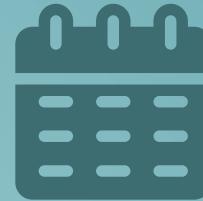
Space and Time Data



14

**Complete Stations
Islandwide**

Recorded as Latitude, Longitude



24

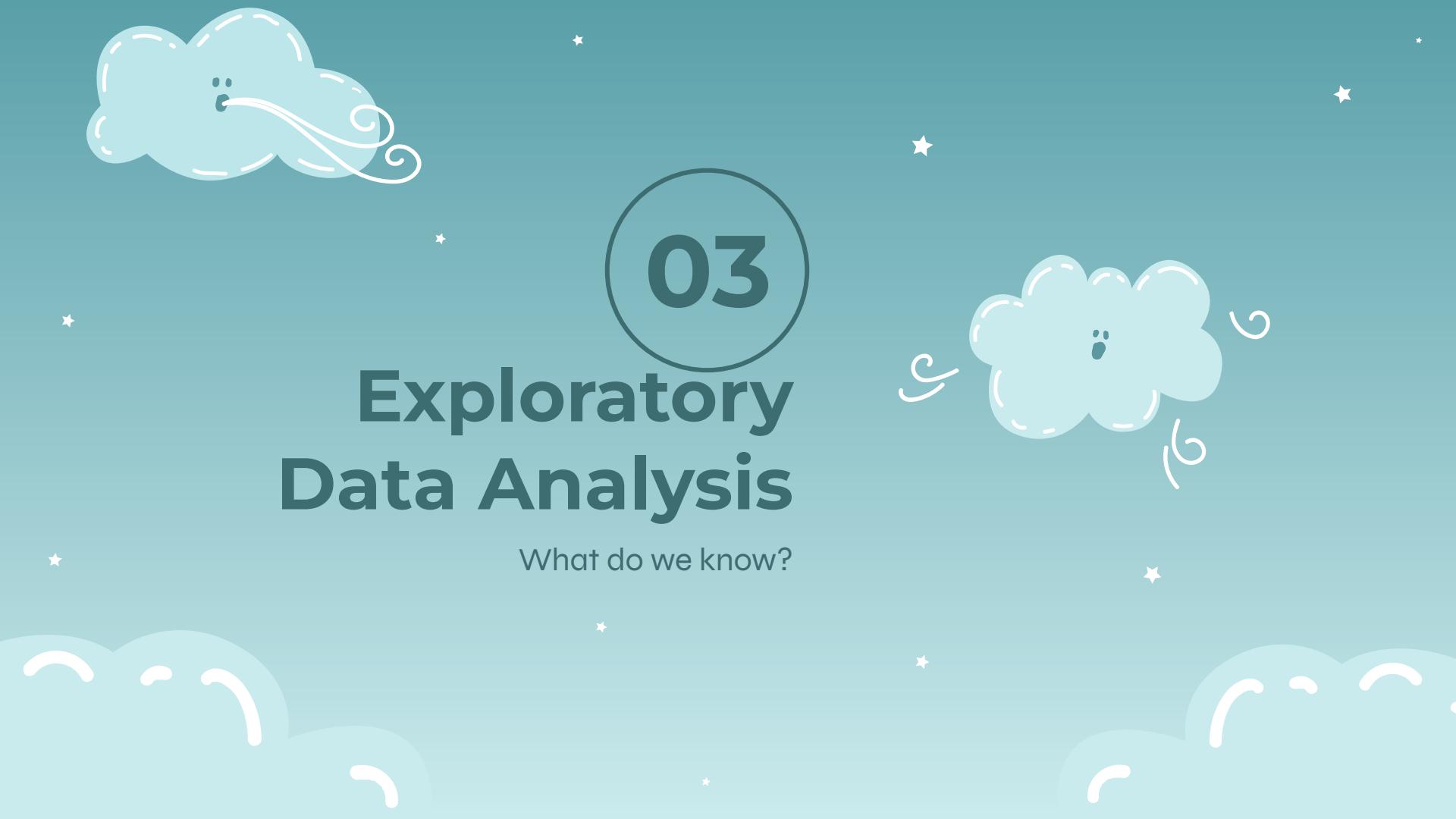
**Complete
Months**

Recorded as YYYY-MM-DD HH:MM:SS

3,178,161

Rows of data collected

(with 9 columns of base features)

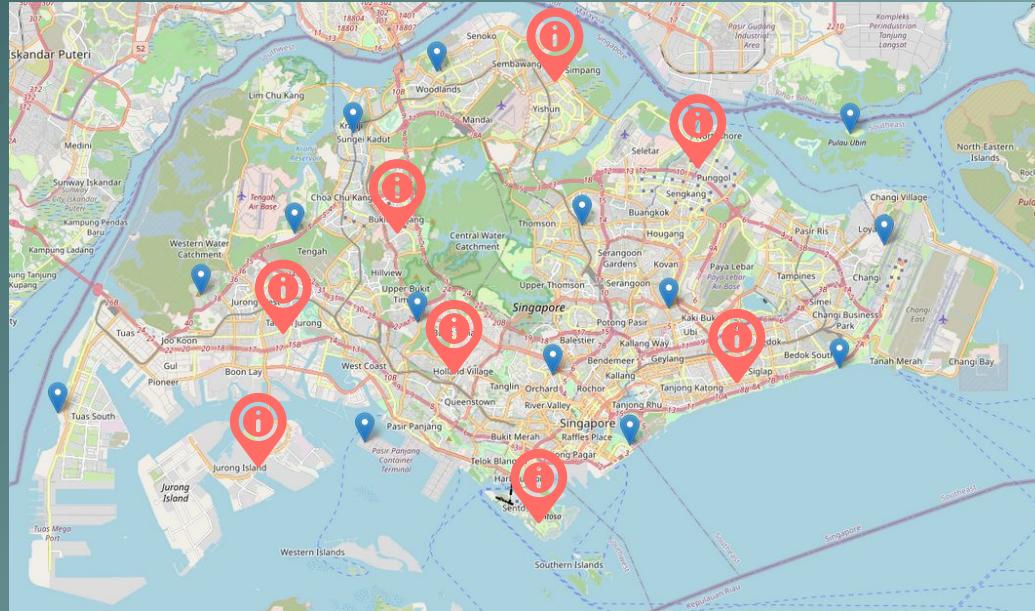


03

Exploratory Data Analysis

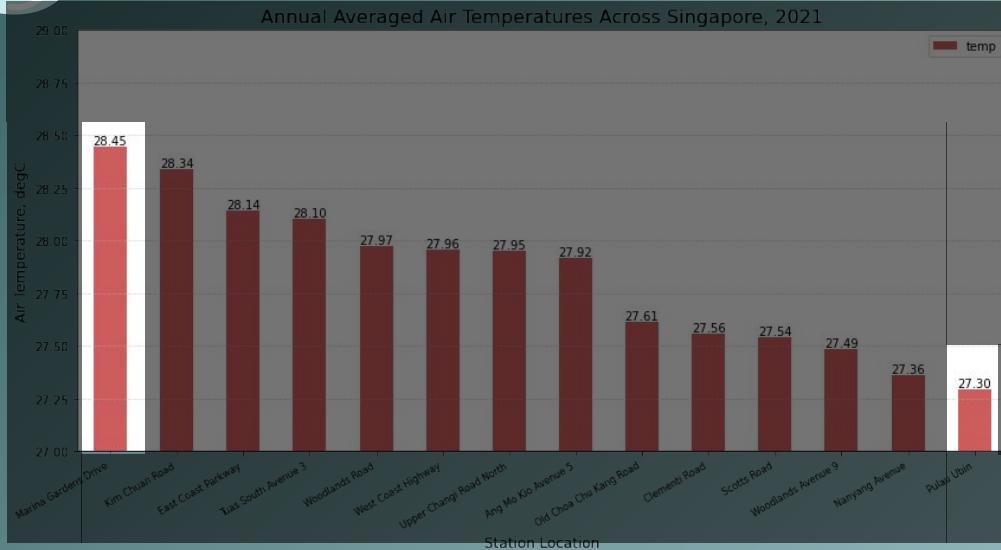
What do we know?

14 Complete Stations Islandwide

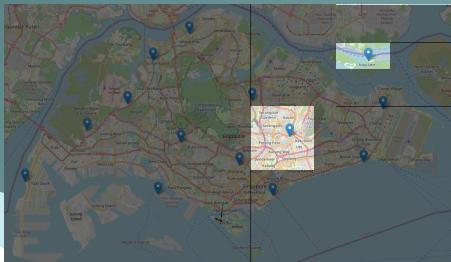


Under
representation in
certain areas

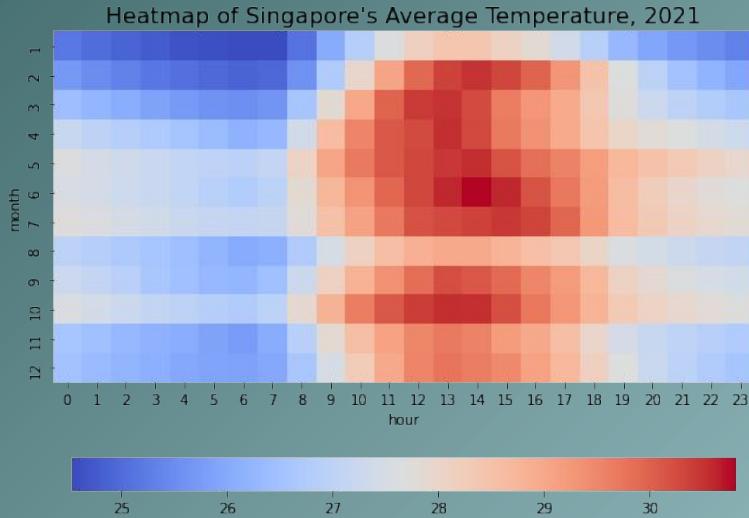
Hotspots Identified - 2021



- **Ang Mo Kio / Tai Seng area observed to be hot**
 - Poor wind penetration inland
- **Pulau Ubin area also observed to be coldest**
 - Rain clouds from Pasir Gudang
 - Dense vegetation

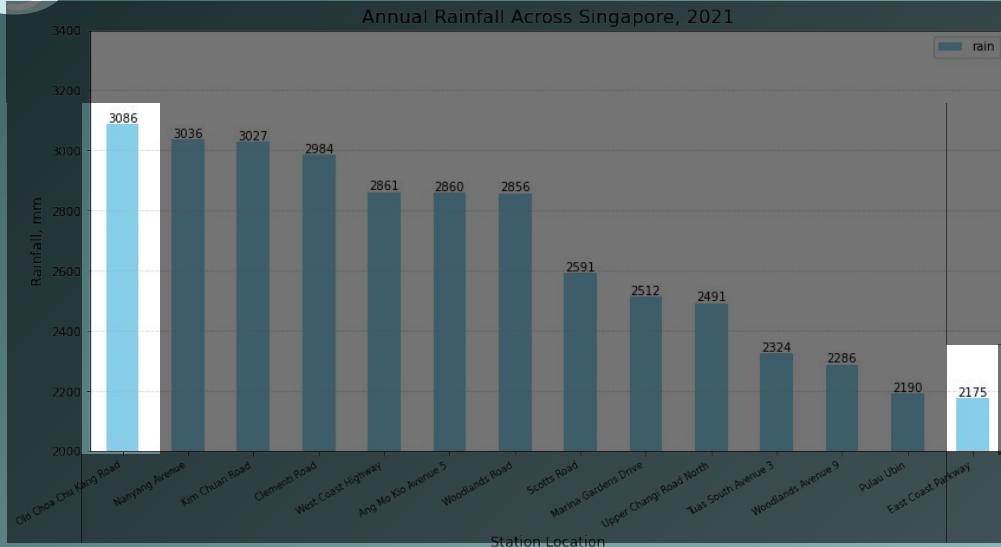


Hotspots Identified - 2021

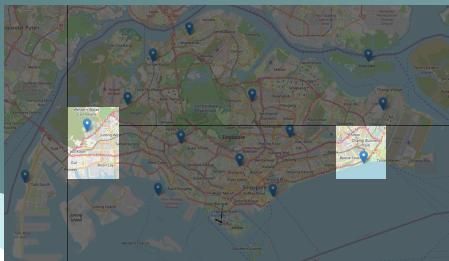


- **Hottest periods**
 - **MONTHS**
 - FEB, MAY, JUNE, OCT
 - **TIME OF DAY**
 - 2-3PM

Rain Areas - 2021

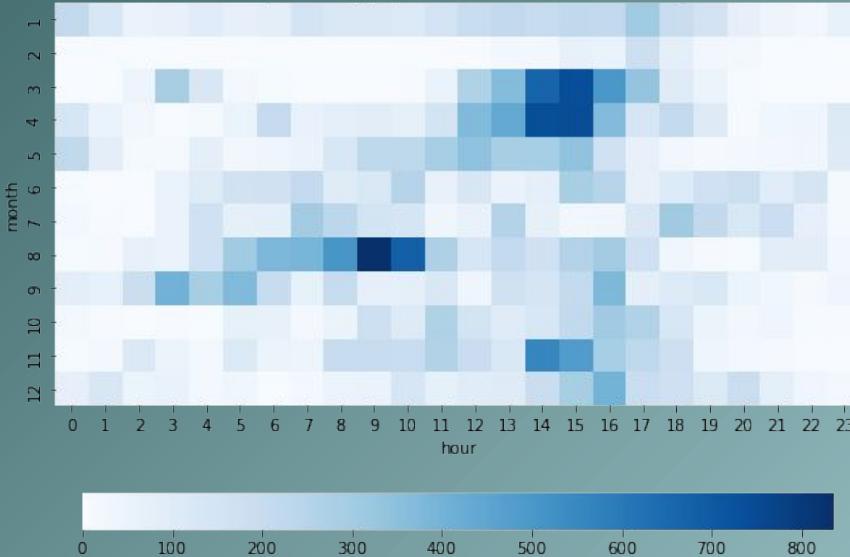


- **Choa Chu Kang area observed to get the most rainfall**
 - Rains from Sumatran Islands
- **East Coast Park area also observed to receive the least**



Rain Areas - 2021

Heatmap of Singapore's Annual Rainfall, 2021



- **Wettest periods**

- **MONTHS**

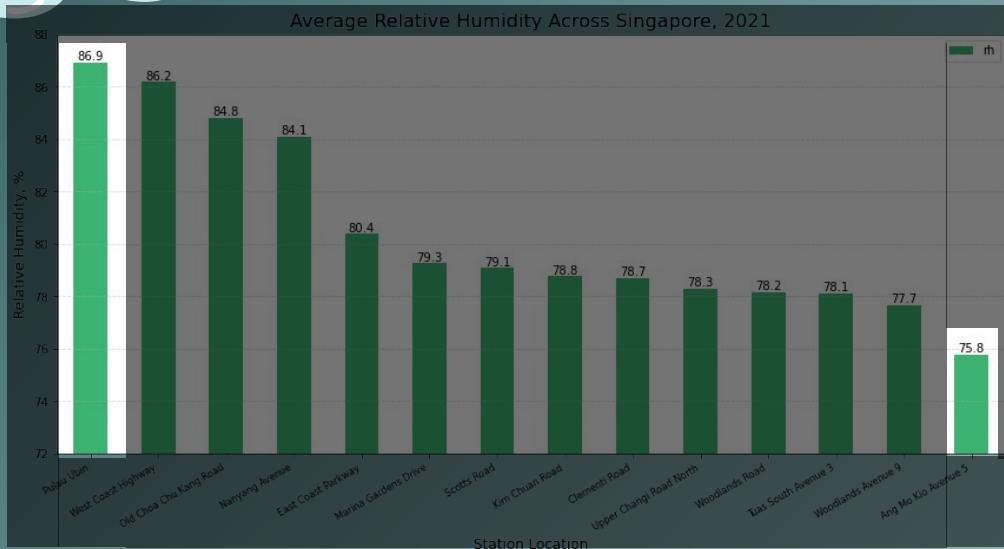
- JAN, MAR, APR, AUG, DEC

- **TIME OF DAY**

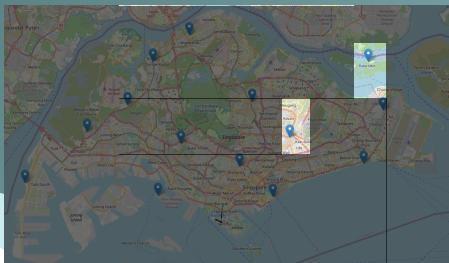
- 1-4PM

- 7-10AM

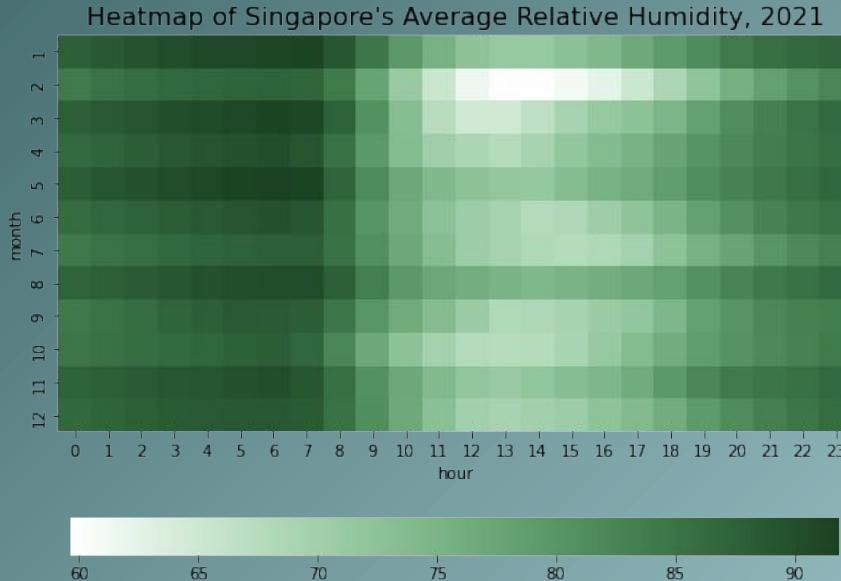
Relative Humidity - 2021



- **Pulau Ubin area observed be the most humid**
 - Dense vegetation
- **Ang Mo Kio East area also observed to be the least humid**

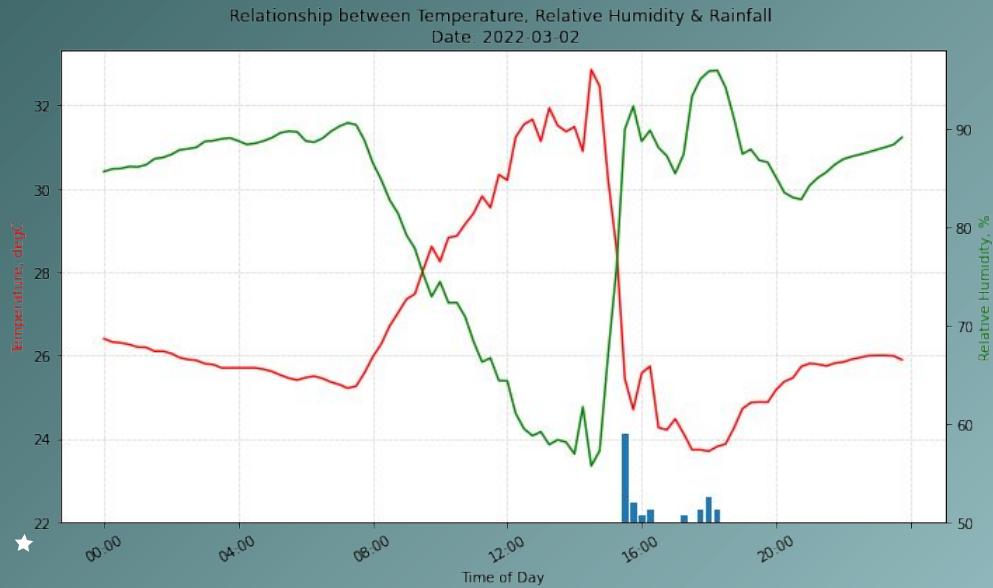


Relative Humidity - 2021

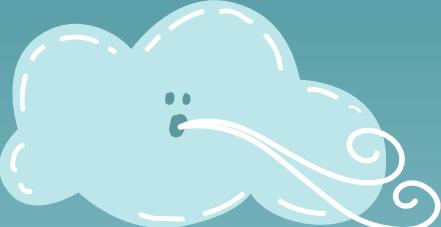


- **Most Humid Periods (day time)**
 - **MONTHS**
 - MAY, JUN, JUL, AUG
 - OCT, NOV
 - **TIME OF DAY**
 - 2-5PM

A Nice Neat Bow...



- Sudden drop in **temperature**
- Sharp surge in **relative humidity**
- **Continued rainfall observed**



04

Feature Engineering

What else can we get?

Features Created



Lag Columns



Pairwise Distance



Pairwise Bearings



Rain Matrix



Interaction
Features

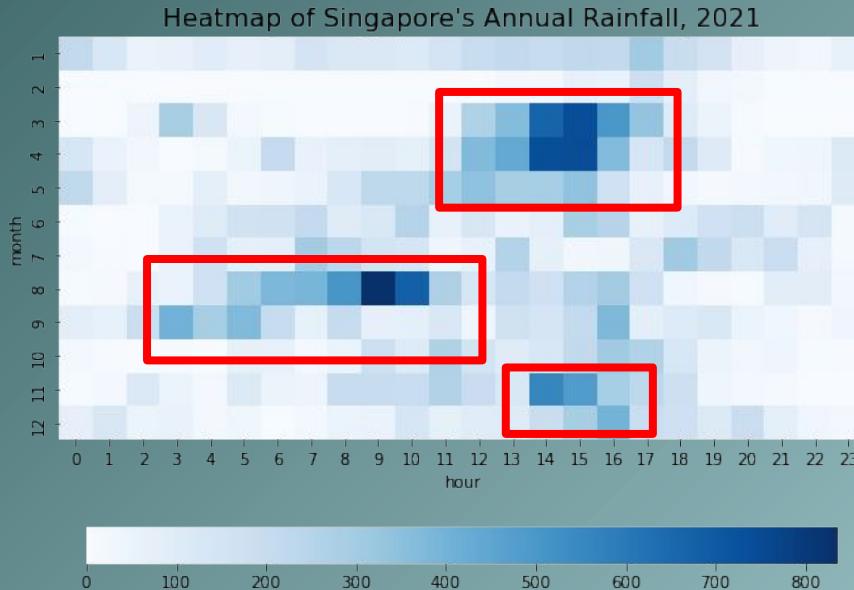


05

Scoring & Baseline

How high is high?

Time-based Rain Occurrence



- **Time-based intuition based on past data**
 - MAR-MAY, 11-6PM
 - AUG-SEP, 3-11AM
 - NOV-DEC 2-5PM

Scoring Metric - Baseline

Confusion Matrix		Predictions	
Actuals	No Rain (0)	Rain (1)	
	No Rain (0)	2560363	338781
	Rain (1)	27598	10261

Custom Scoring

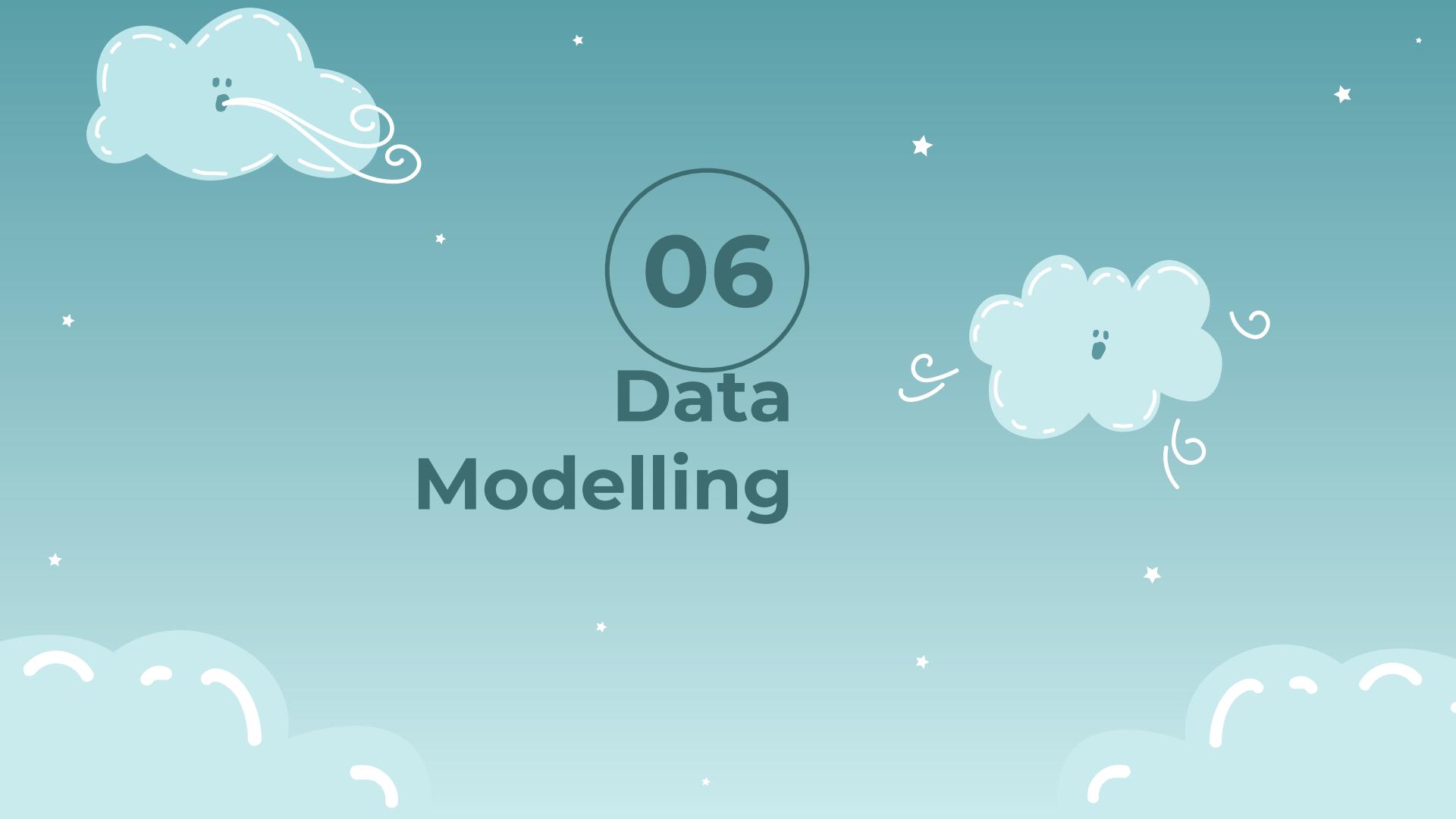
$$= 0.75 * \text{Recall} + 0.25 * \text{Precision}$$

$$= 21.06\%$$

Precision = 2.94%

Recall = 27.10%

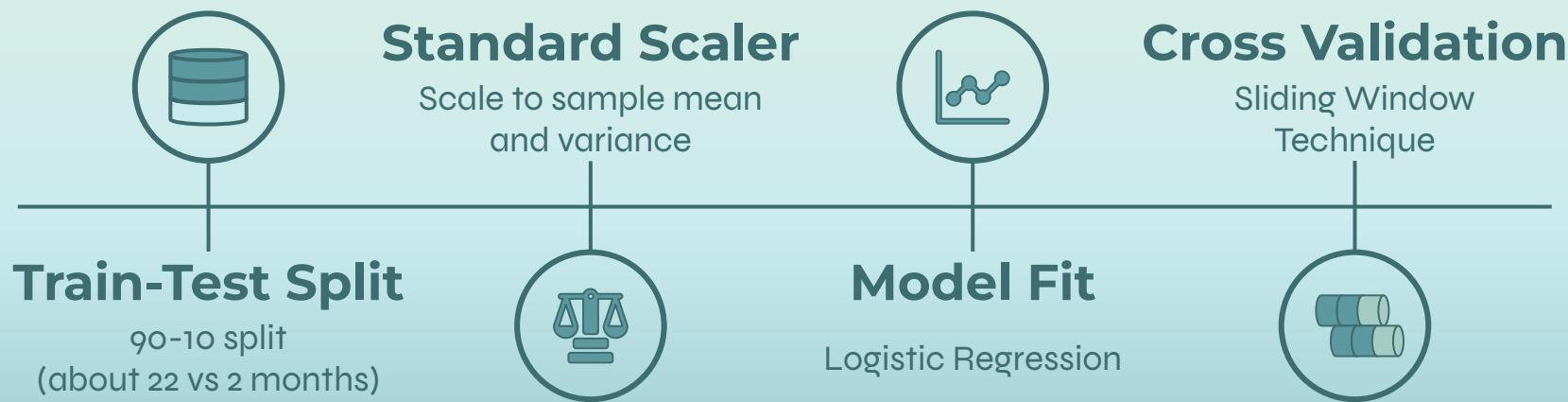
- Extremely conservative baseline approach with high false positive count
- Balanced Custom Scoring where Recall score is worth 3x more than Precision



06

Data Modelling

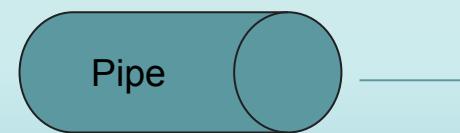
Modelling Pipeline - Station Level



Modelling Pipeline - Scoring



**Station Level
Data**



**Modelling
Pipeline**

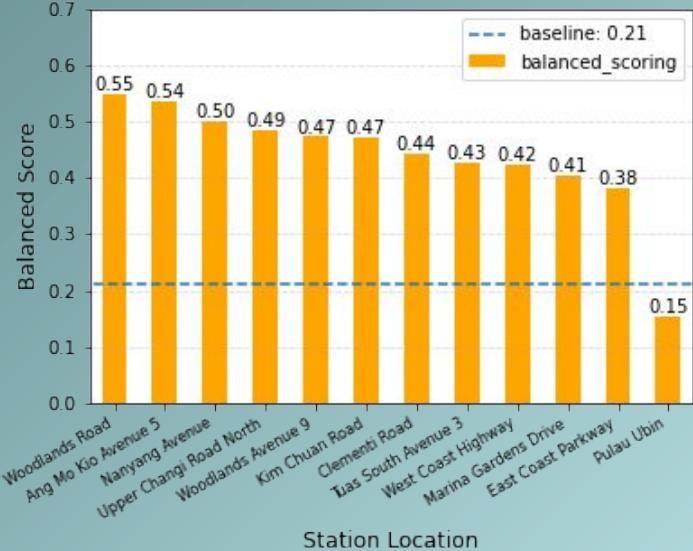


**Generate
Predictions**

Modelling Results - Balanced Scoring

	Station	tn	tp	fn	fp	balanced_scoring	name
0	station_S100	20825	107	93	75	0.548228	Woodlands Road
5	station_S109	19249	142	145	71	0.537747	Ang Mo Kio Avenue 5
10	station_S44	21069	106	131	53	0.502110	Nanyang Avenue
8	station_S24	21105	92	110	68	0.485334	Upper Changi Road North
1	station_S104	19361	117	165	63	0.473670	Woodlands Avenue 9
9	station_S43	19681	117	180	50	0.470604	Kim Chuan Road
11	station_S50	16337	77	117	54	0.444627	Clementi Road
6	station_S115	20620	55	66	102	0.428489	Tuas South Avenue 3
7	station_S116	19459	77	133	52	0.424225	West Coast Highway
4	station_S108	15692	66	141	33	0.405797	Marina Gardens Drive
3	station_S107	20353	69	157	44	0.381637	East Coast Parkway
2	station_S106	19561	16	156	32	0.153101	Pulau Ubin

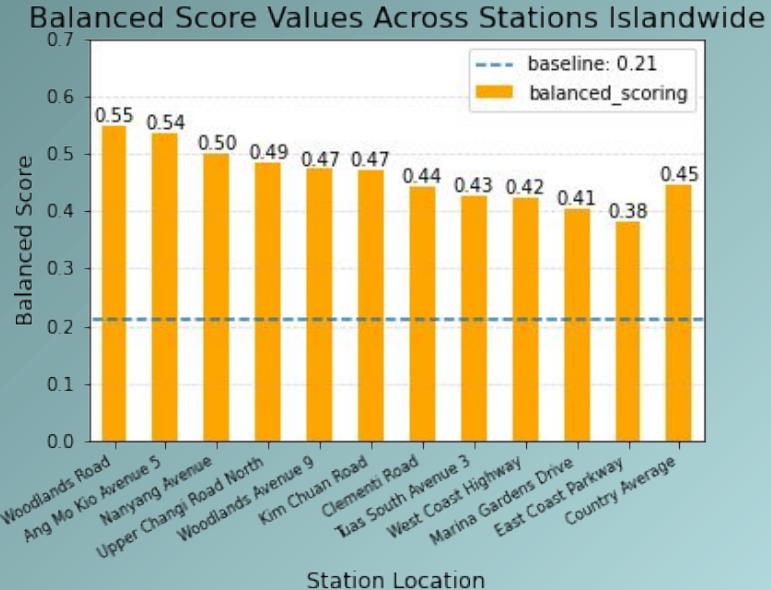
Balanced Score Values Across Stations Islandwide



- With the exception of Pulau Ubin, all stations fared better than the baseline score

Modelling Results - Balanced Scoring

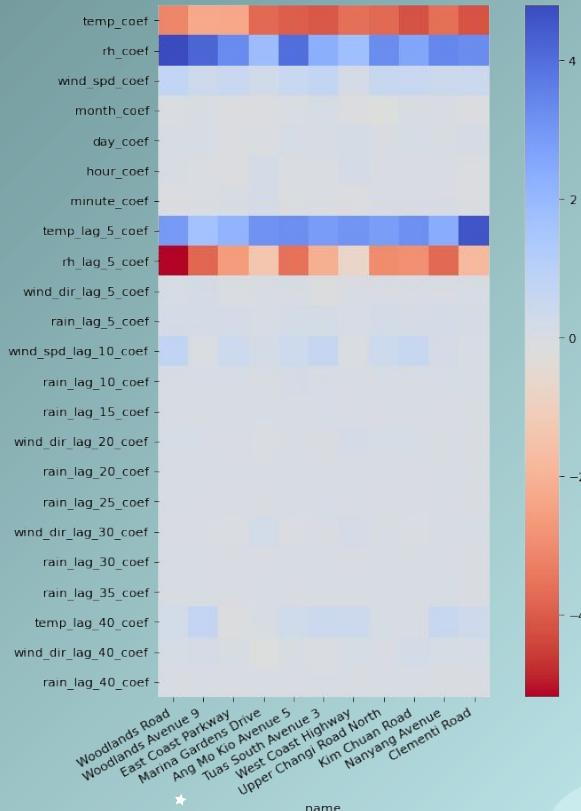
	Station	tn	tp	fn	fp	balanced_scoring	name
0	station_S100	20825	107	93	75	0.548228	Woodlands Road
5	station_S109	19249	142	145	71	0.537747	Ang Mo Kio Avenue 5
10	station_S44	21069	106	131	53	0.502110	Nanyang Avenue
8	station_S24	21105	92	110	68	0.485334	Upper Changi Road North
1	station_S104	19361	117	165	63	0.473670	Woodlands Avenue 9
9	station_S43	19681	117	180	50	0.470604	Kim Chuan Road
11	station_S50	16337	77	117	54	0.444627	Clementi Road
6	station_S115	20620	55	66	102	0.428489	Tuas South Avenue 3
7	station_S116	19459	77	133	52	0.424225	West Coast Highway
4	station_S108	15692	66	141	33	0.405797	Marina Gardens Drive
3	station_S107	20353	69	157	44	0.381637	East Coast Parkway
2	station_S106	19561	16	156	32	0.153101	Pulau Ubin



- While low, Country-averaged model performed almost 2x better than baseline score

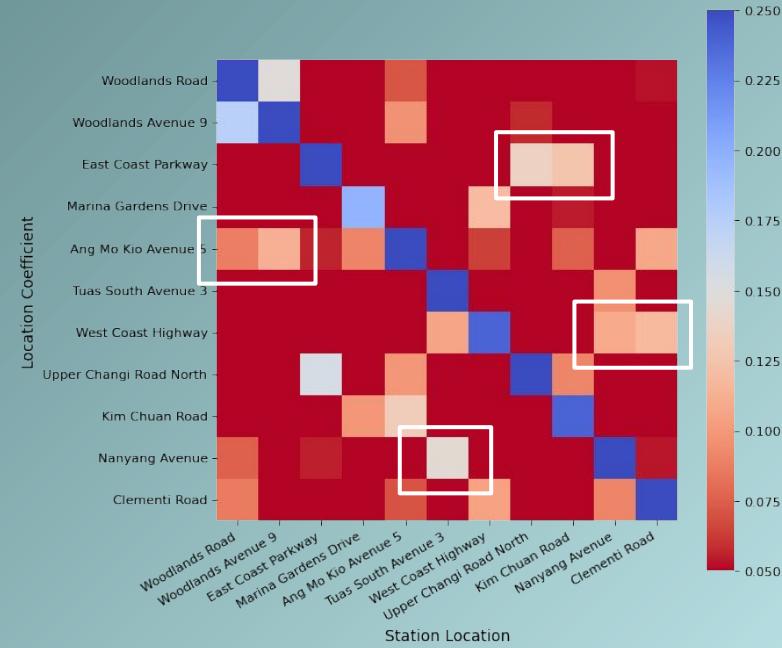
Modelling Results - Top Contributors

- Clearly, temperature and relative humidity played the biggest role in the prediction
 - The inverse trend is observed with the lagged values, up to 5min lookback
 - Minor contribution from wind speed values



Modelling Results - Top Contributors

- Certain locations appear to have a rain predictor relationship
 - AMK TO WOODLANDS
 - EAST COAST PARK TO UPPER CHANGI
 - WEST COAST TO JURONG WEST / CLEMENTI
 - JURONG WEST AND TUAS

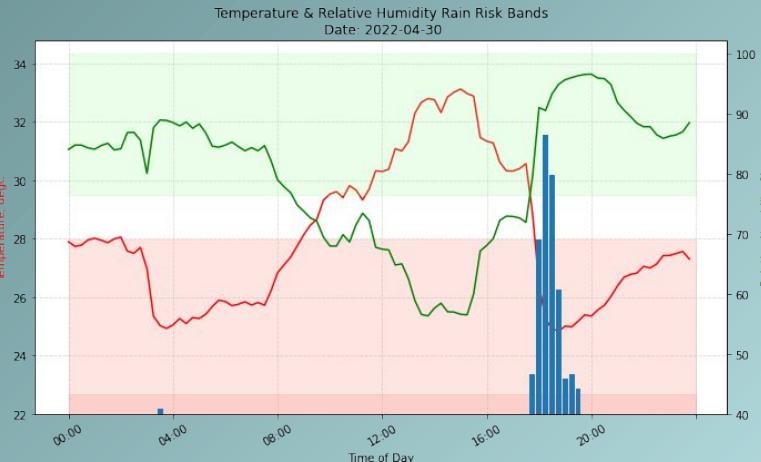


Modelling Results - Coefficient Interpretation

	Station	name	temp_mean	temp_vars	temp_coef	temp_odds	rh_mean	rh_vars	rh_coef	rh_odds
0	station_S100	Woodlands Road	28.0063	6.1391	-3.1387	23.073851	78.4199	154.0241	4.7971	121.158548
1	station_S104	Woodlands Avenue 9	27.6044	5.0946	-2.2973	9.947288	77.6331	131.1602	4.1986	66.593035
2	station_S106	Pulau Ubin	27.2947	5.9110	-3.9441	51.629850	86.2777	149.0348	2.1227	8.353662
3	station_S107	East Coast Parkway	28.1139	2.4001	-2.3206	10.181782	80.8486	59.1088	3.3068	27.297633
4	station_S108	Marina Gardens Drive	28.9044	4.2299	-3.7506	42.546602	79.7949	161.9771	1.8622	6.437885
5	station_S109	Ang Mo Kio Avenue 5	27.9699	5.2910	-3.9699	52.979233	76.4757	134.2866	4.0134	55.334689
6	station_S115	Tuas South Avenue 3	28.0936	3.4935	-4.0875	59.590729	78.8653	74.6690	2.2737	9.715281
7	station_S116	West Coast Highway	28.1666	2.7078	-3.6060	36.818484	86.0655	75.3240	1.7173	5.569471
8	station_S24	Upper Changi Road North	27.9289	4.2627	-3.7587	42.892629	78.9099	99.6138	3.2706	26.327131
9	station_S43	Kim Chuan Road	28.4834	3.8542	-4.2002	66.699670	79.6486	119.3080	2.6041	13.519053
10	station_S44	Nanyang Avenue	27.3712	4.9219	-3.5859	36.085820	82.9914	141.3675	3.4285	30.768763
11	station_S50	Clementi Road	27.8103	4.5567	-4.1431	62.997812	79.5897	128.1876	3.2531	25.870414

```
=====
Report =====
Temp Avg Range:      27.3 - 28.9 degC
Temp Var Range:      2.4 - 6.1 degC
Temp Odds Range:     9 - 66

RH Avg Range:        76.5 - 86.3 %
RH Var Range:        59.1 - 162.0 %
RH Odds Range:       5 - 121
=====
End Report =====
```



As temperature drops by ~3 degC, odds of rain continuing for the next 20min increase by over 50 times.



07

Conclusions

Concluding Remarks

Model Wins

- Model was successful at delivering an improved balanced scoring result
- Only 22 months of training data was used
- Model's can be easily interpretable, with temperature and RH as main drivers

Improvement Areas

- Class balancing / SMOTE
- Tuning LogReg threshold value
- Cartesian coordinate pairwise matrix
- Obtain interaction terms





THANKS!

CREDITS: This presentation template was created by [Slidesgo](#),
including icons by [Flaticon](#) and infographics & images by [Freepik](#)

Please keep this slide for attribution

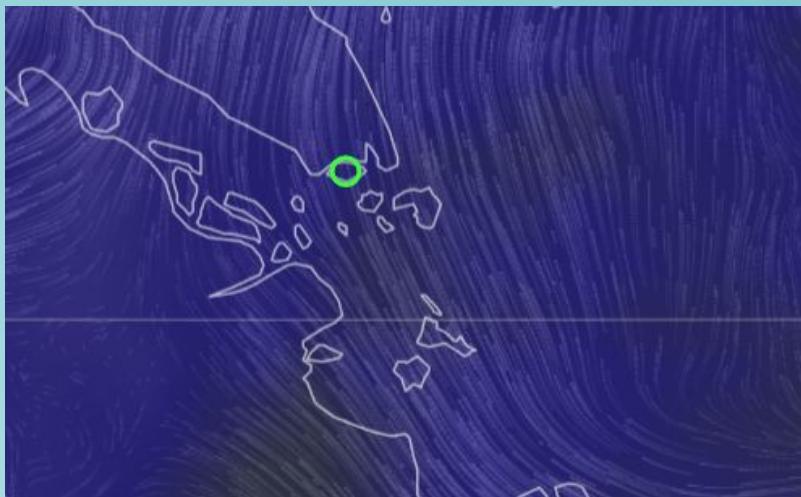
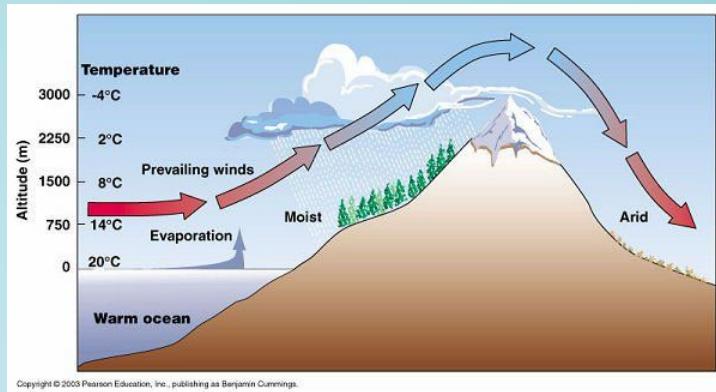


Problems Faced - Pairwise Interactions

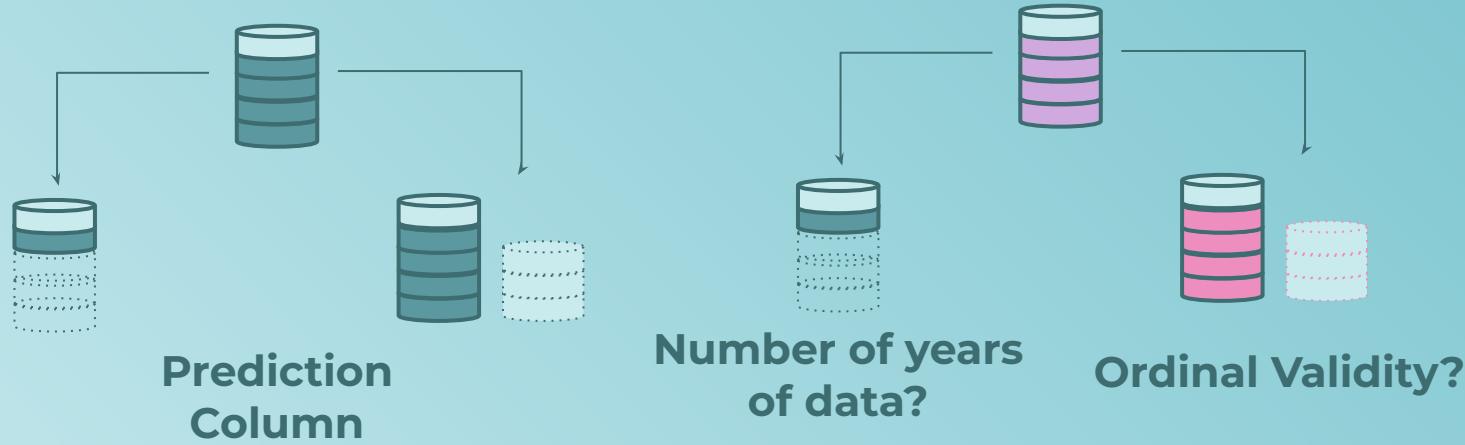


System	Metric	Radial	Cartesian
Others	Distance	✓	✓
	Bearing	✓	-
Self	Distance	✓	✓
	Bearing	✗	-

Wind Data

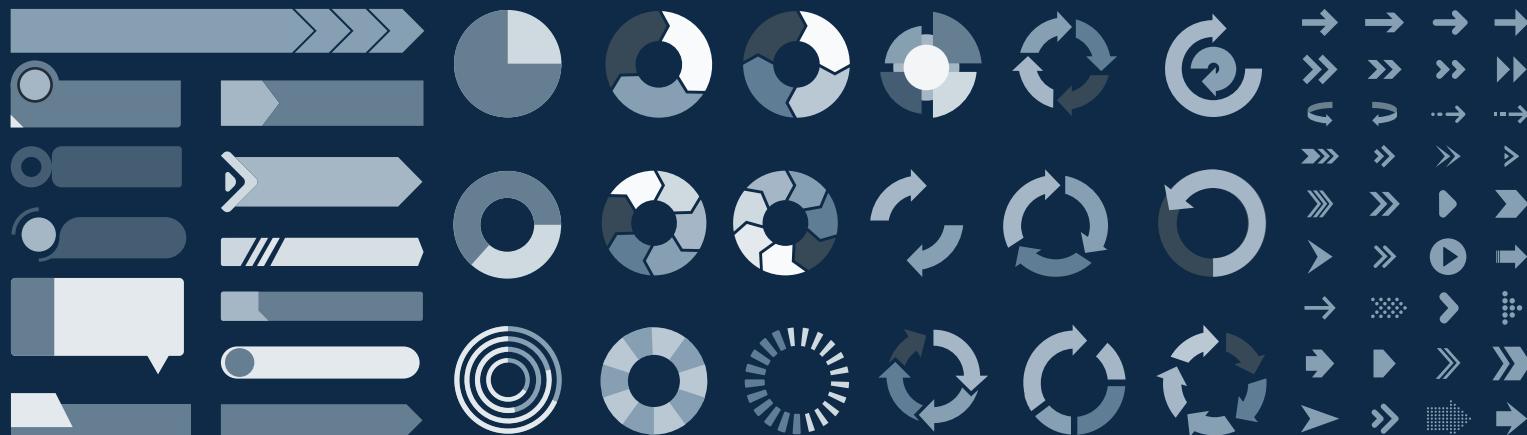


Time Series SMOTE Complications

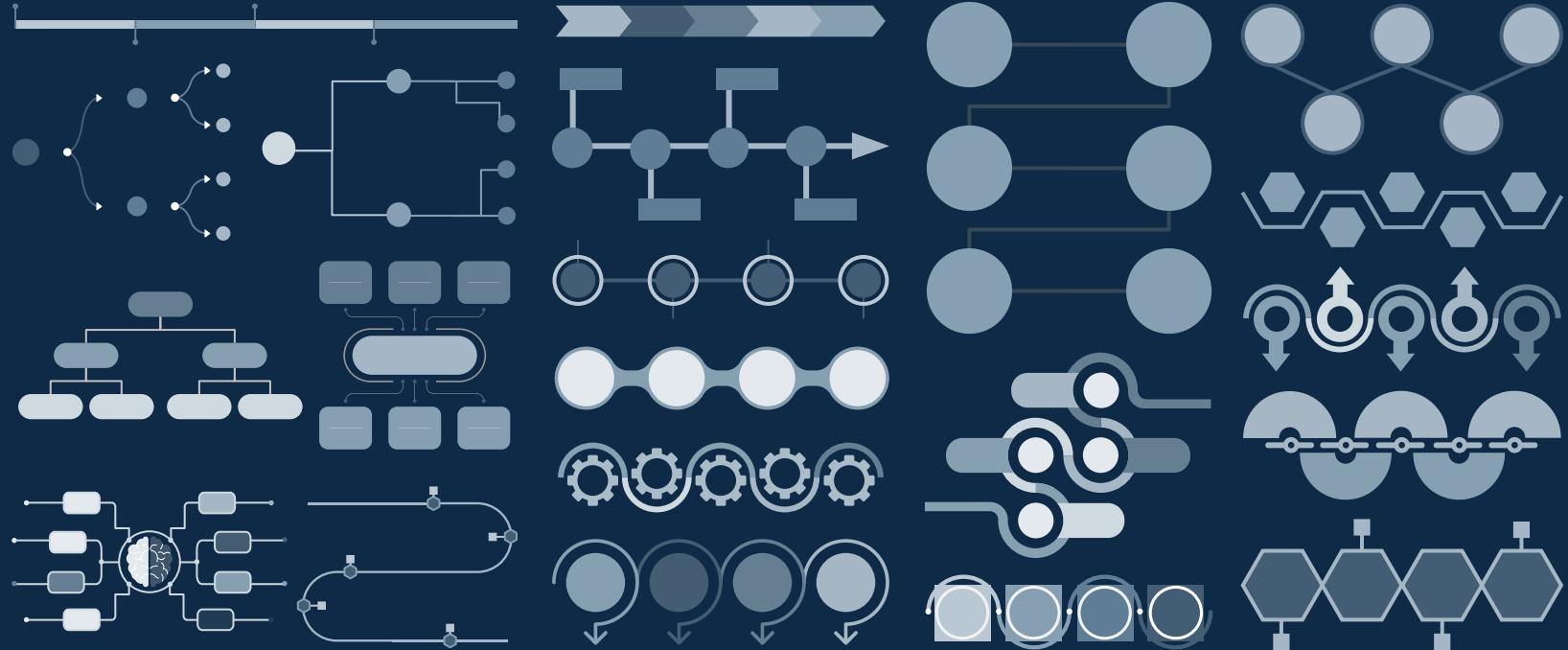


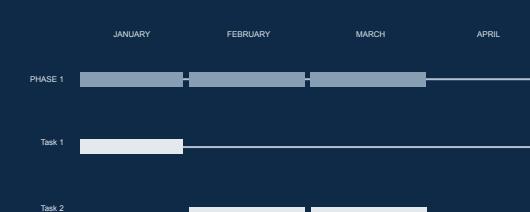
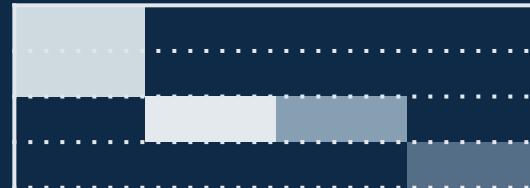
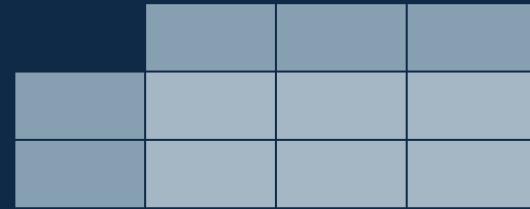
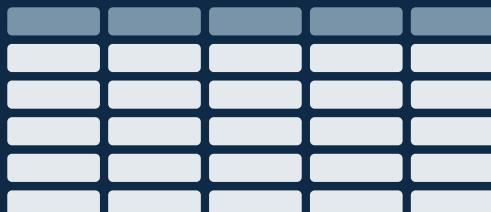
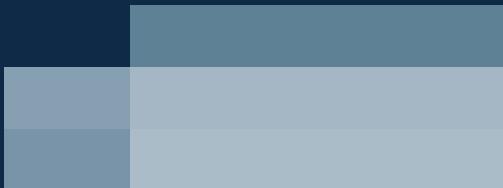
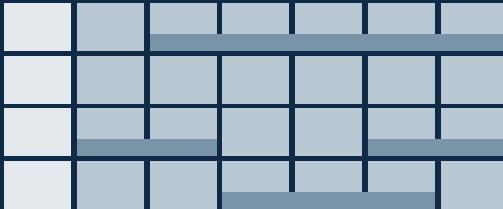
Use our editable graphic resources...

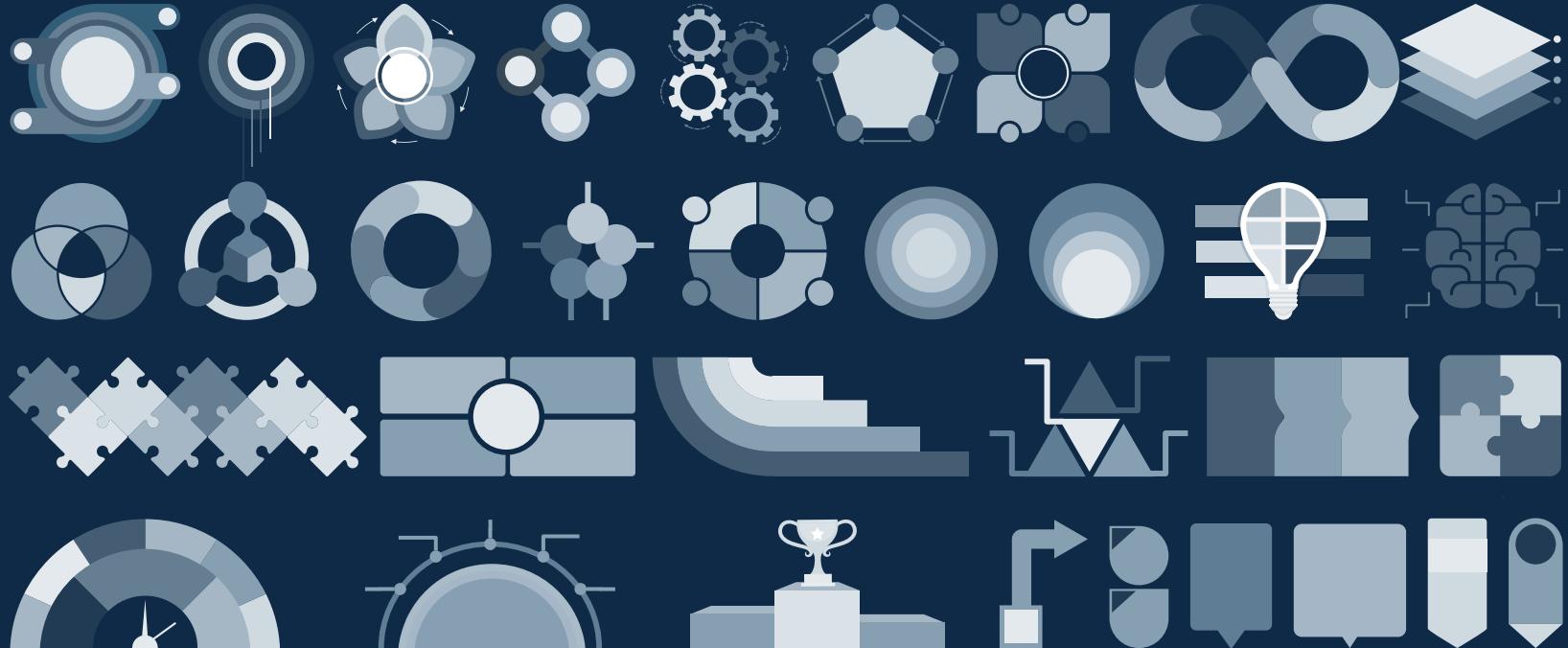
You can easily resize these resources without losing quality. To change the color, just ungroup the resource and click on the object you want to change. Then, click on the paint bucket and select the color you want. Group the resource again when you're done. You can also look for more infographics on Slidesgo.

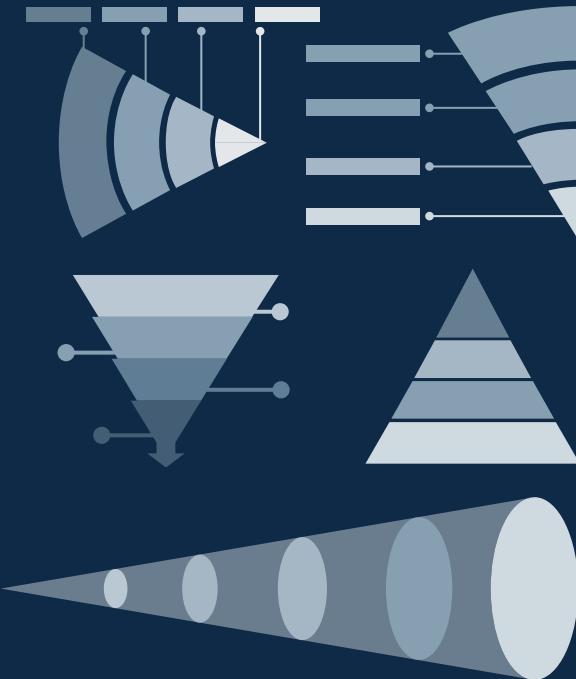
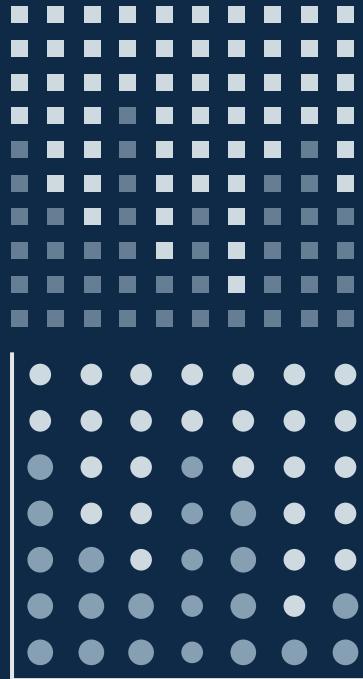












...and our sets of editable icons

You can resize these icons without losing quality.

You can change the stroke and fill color; just select the icon and click on the paint bucket/pen.

In Google Slides, you can also use Flaticon's extension, allowing you to customize and add even more icons.



Educational Icons



Medical Icons



Business Icons



Teamwork Icons



Help & Support Icons



Avatar Icons



Creative Process Icons



Performing Arts Icons



Nature Icons



SEO & Marketing Icons



