

West Nile Virus Prediction

Group 2 - Chee Tzen, Zhi Yuan,
Raymond, Mitchelle & Kris

Table of contents

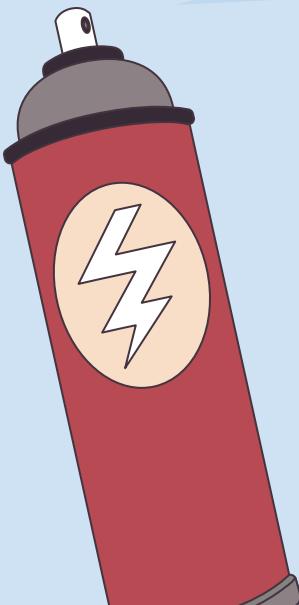
01 Problem Statement

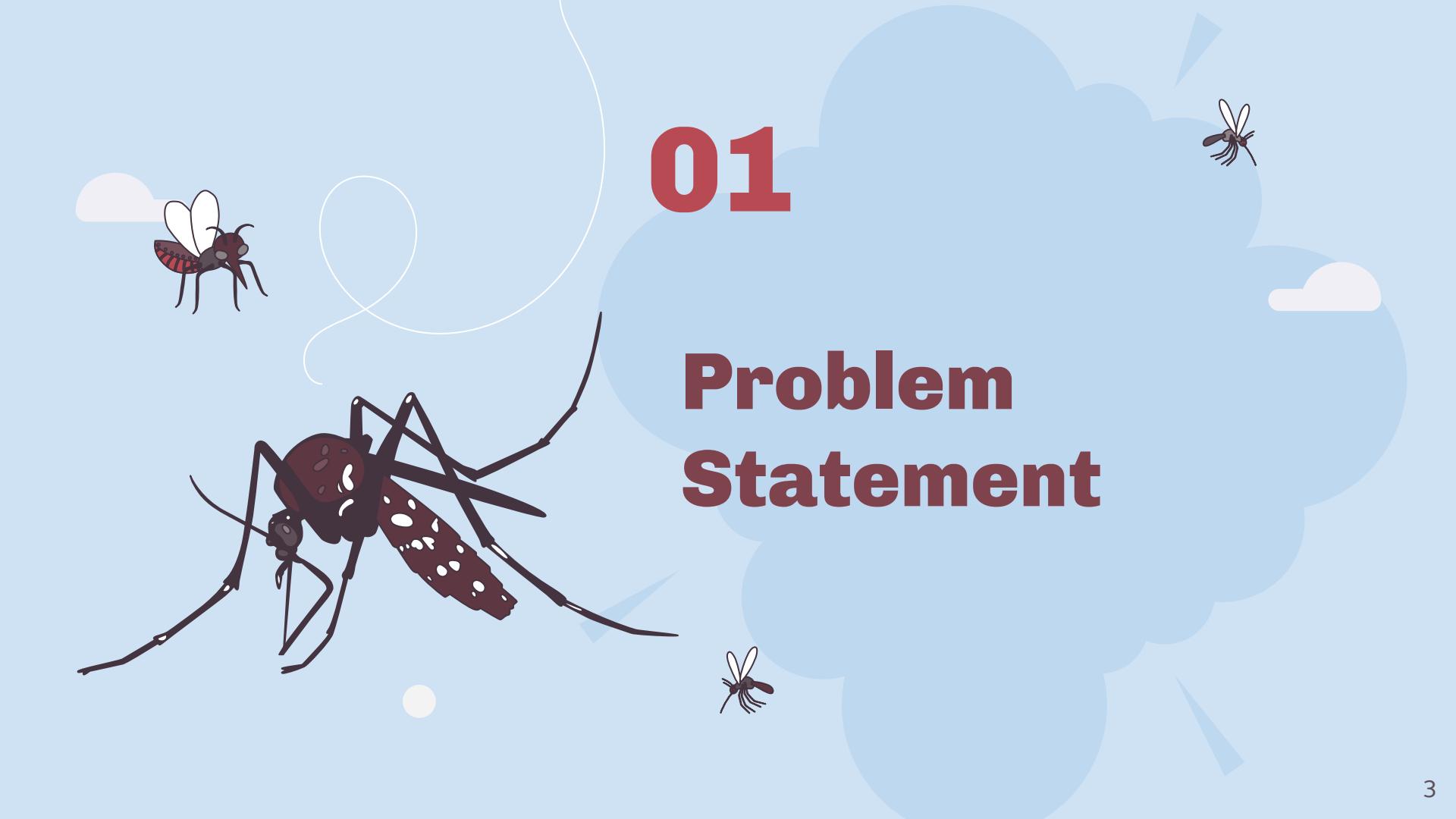
**03 Preprocessing,
Modelling &
Evaluation**

**02 Data Cleaning &
EDA**

**04 Cost Benefit
Analysis**

**05 Conclusion &
Recommendation**





01

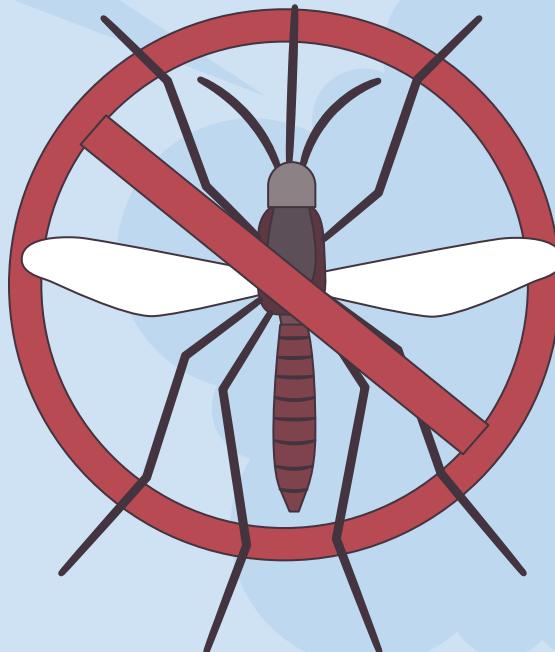
Problem Statement

West Nile Virus (WNV) is the leading mosquito-borne disease in the United States¹ which also contains the Zika, dengue and yellow fever virus. Its primary hosts are birds that mosquitoes feed on which in turn spread to people and other animals through their bites.

While WNV is not extremely virulent, there is currently no vaccine against it.



1. <https://www.cdc.gov/westnile/index.html>



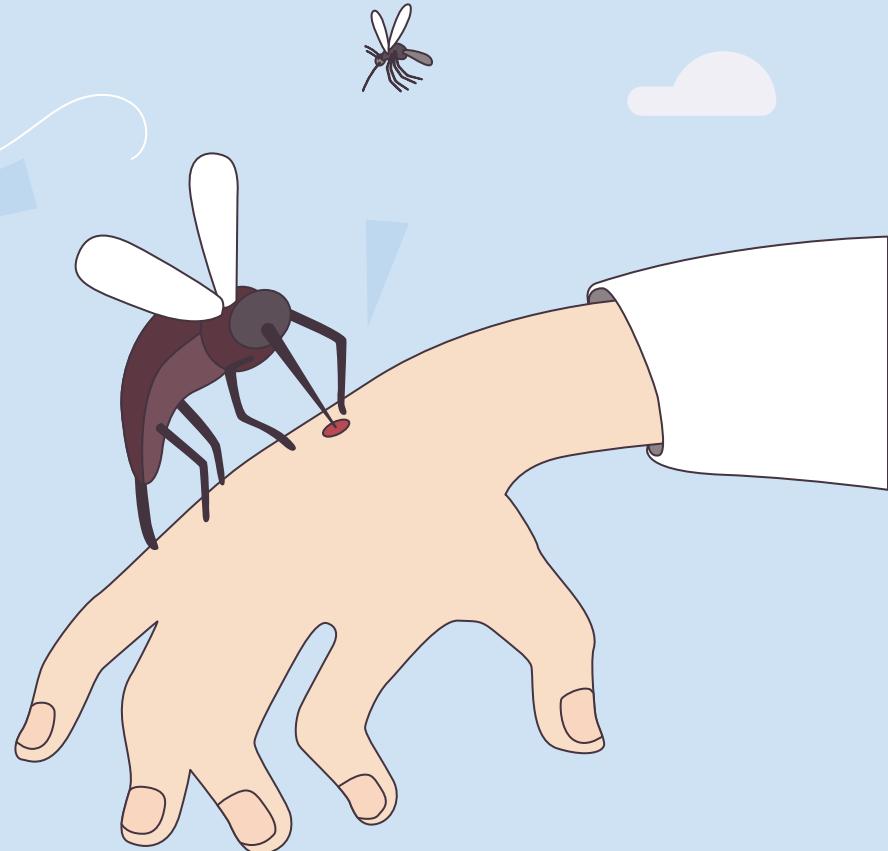
Due to the recent epidemic of WNV in the Windy City, the Department of Public Health has set up a surveillance and control system.

As new data scientists working at the Disease and Treatment Agency, we have been tasked to use the information collected to come up with an effective plan that details when and where to deploy costly yet necessary pesticides throughout the city.



02

Data Cleaning & EDA

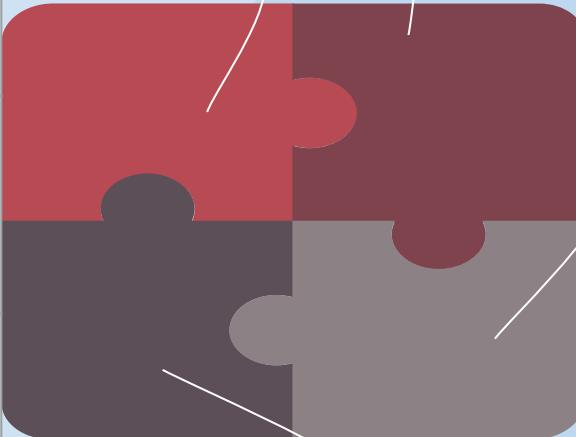


Weather

Columns	Action
<ul style="list-style-type: none">• 'sunrise'• 'sunset'	No data from station 2 so impute from station 1
<ul style="list-style-type: none">• 'Water1'• 'Codesum'• 'snowfall'• 'depth'	0 or missing so drop
<ul style="list-style-type: none">• 'Preciptotal'• 'Avgspeed'• 'Sealevel'• 'stnpressure'	Impute with median values



Data Cleaning



Train and test

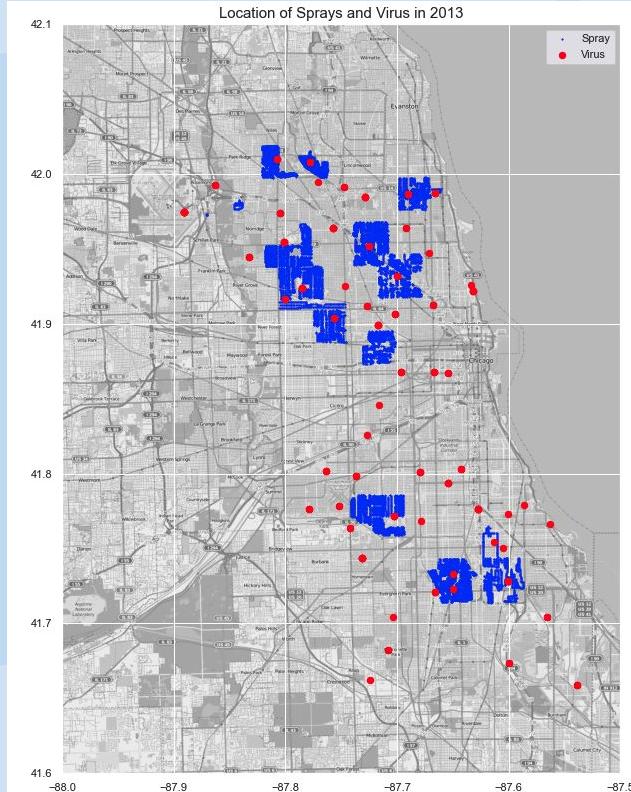
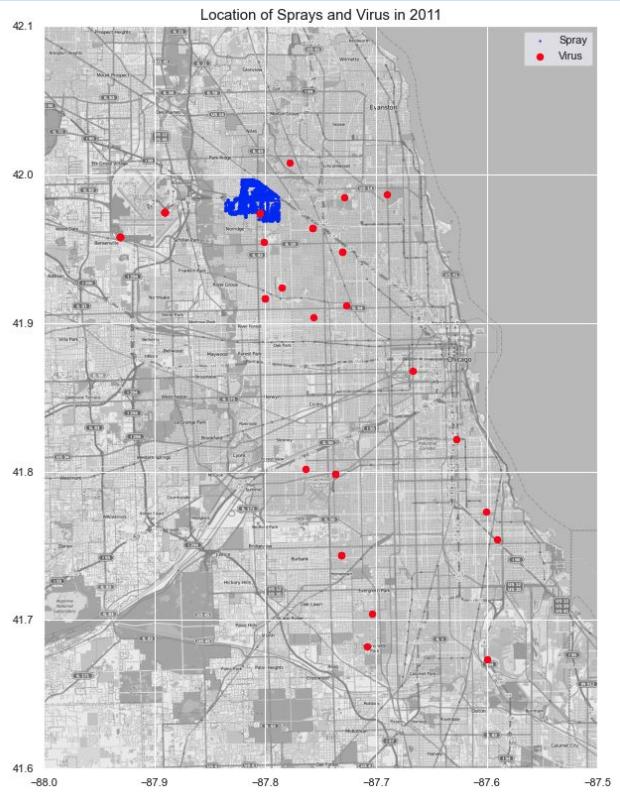
Description: Results organized in such a way that when the number of mosquitoes 50 will form new row
Action: Merge date duplicated rows



Spray

Action: 3 columns of day, month and year in case of analysis required

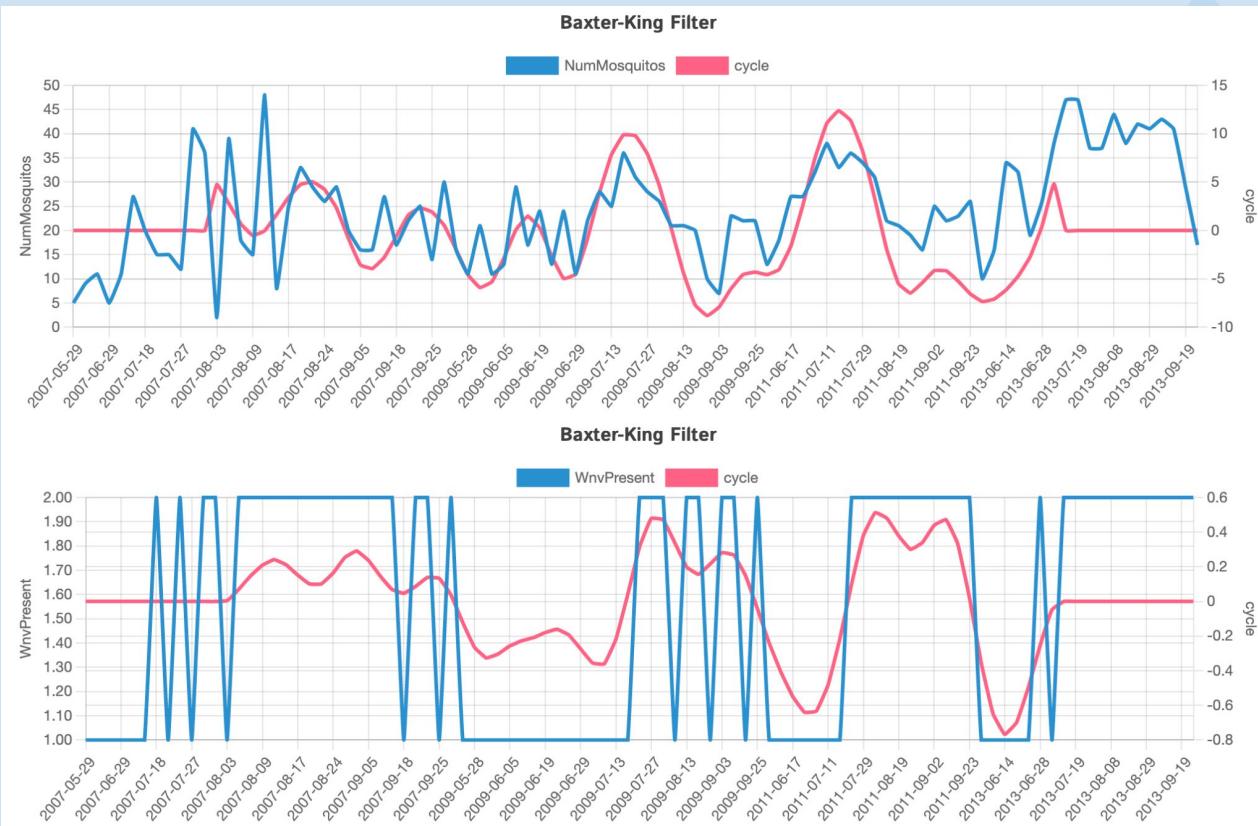
EDA



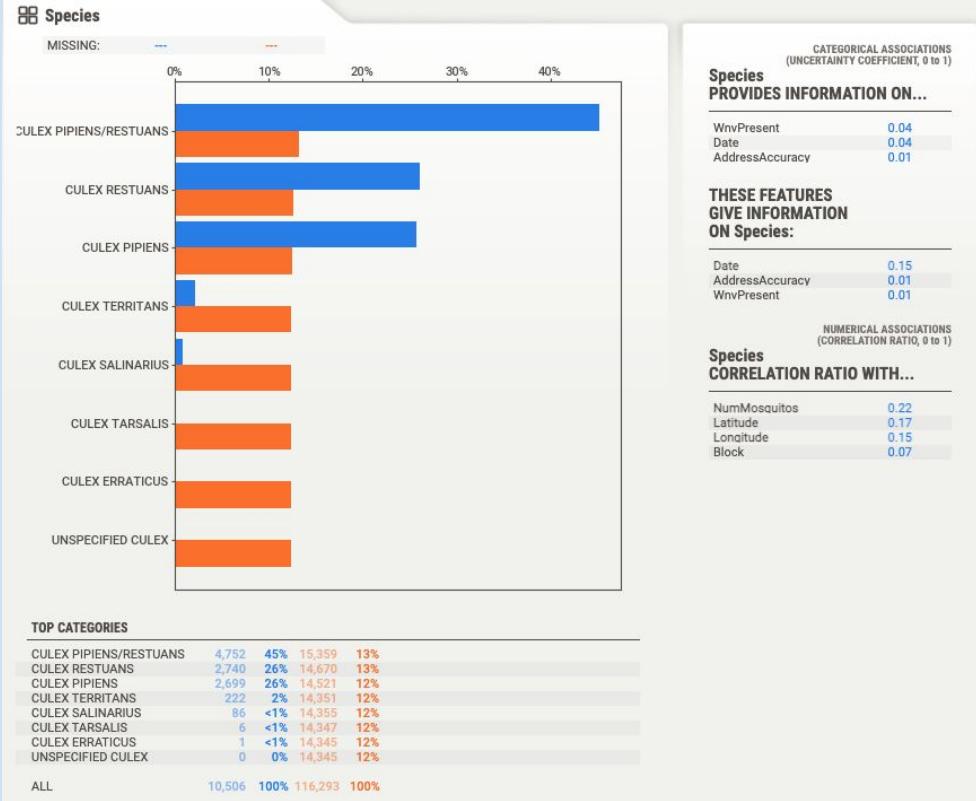
Train vs Test

Training Data		Test Data	
10506	ROWS	116293	
813	DUPLICATES	0	
5.4 MB	RAM	58.7 MB	
12	FEATURES	11	
4	CATEGORICAL	3	
4	NUMERICAL	3	
4	TEXT	4	
ASSOCIATIONS		ASSOCIATIONS	
Training Data		Test Data	
1 (IGNORED: NOT IN SOURCE DATASET)			

Time Series



Species

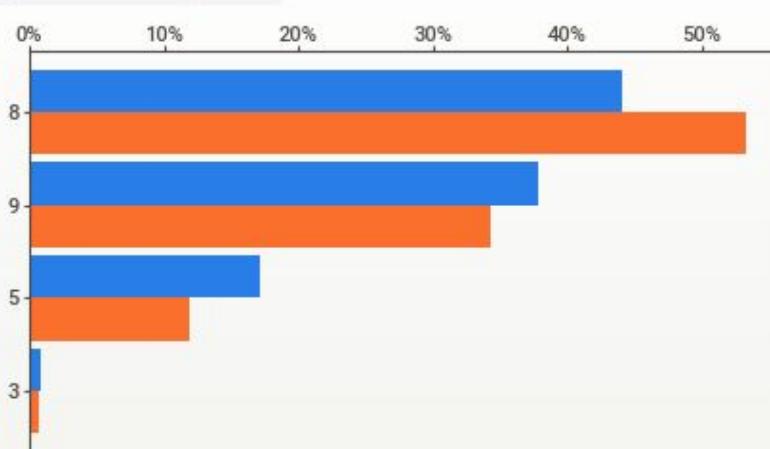


Address Accuracies

~ ~

AddressAccuracy

MISSING:



TOP CATEGORIES

8	4,628	44%	61,973	53%
9	3,980	38%	39,795	34%
5	1,807	17%	13,761	12%
3	91	<1%	764	<1%
ALL	10,506	100%	116,293	100%

CATEGORICAL ASSOCIATIONS
(UNCERTAINTY COEFFICIENT, 0 to 1)

AddressAccuracy PROVIDES INFORMATION ON...

Date	0.03
Species	0.01
WnvPresent	0.00

THESE FEATURES GIVE INFORMATION ON AddressAccuracy:

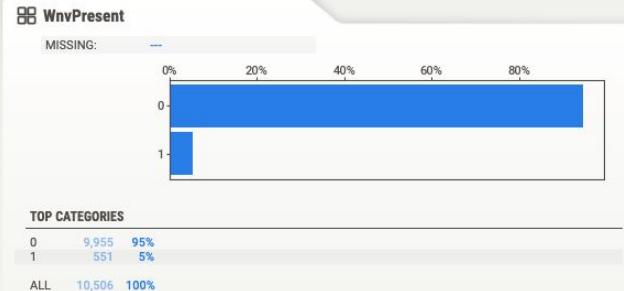
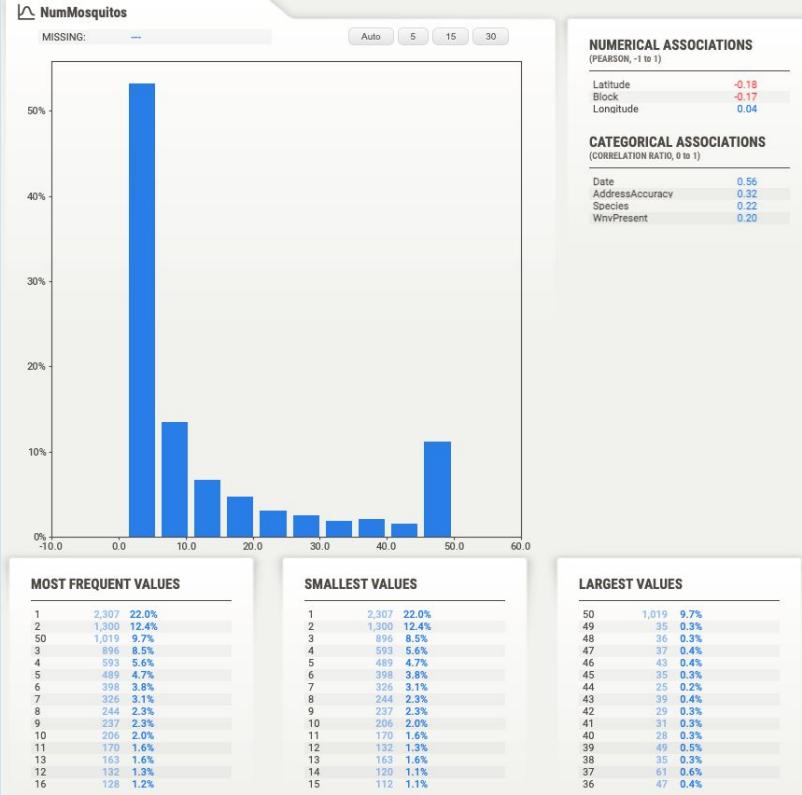
Date	0.11
Species	0.01
WnvPresent	0.00

NUMERICAL ASSOCIATIONS
(CORRELATION RATIO, 0 to 1)

AddressAccuracy CORRELATION RATIO WITH...

Longitude	0.48
Latitude	0.48
Block	0.32
NumMosquitos	0.32

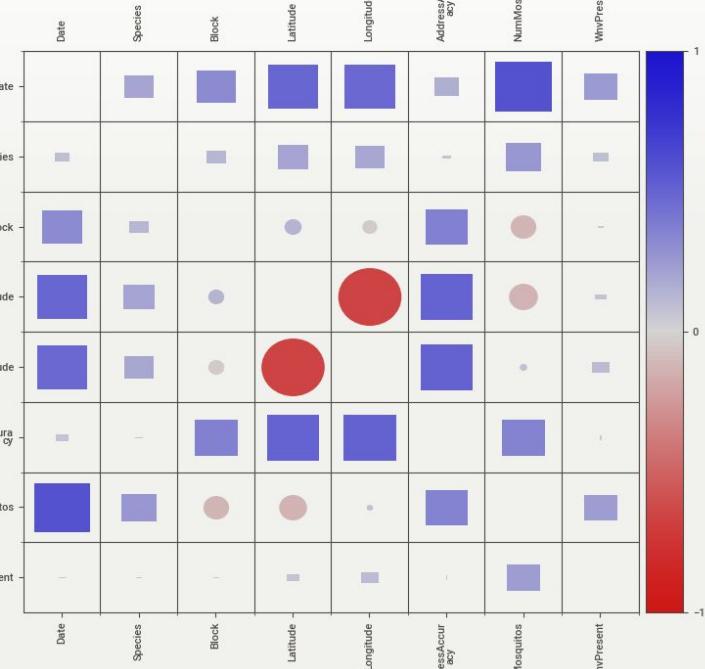
Response



Train vs Test Associations

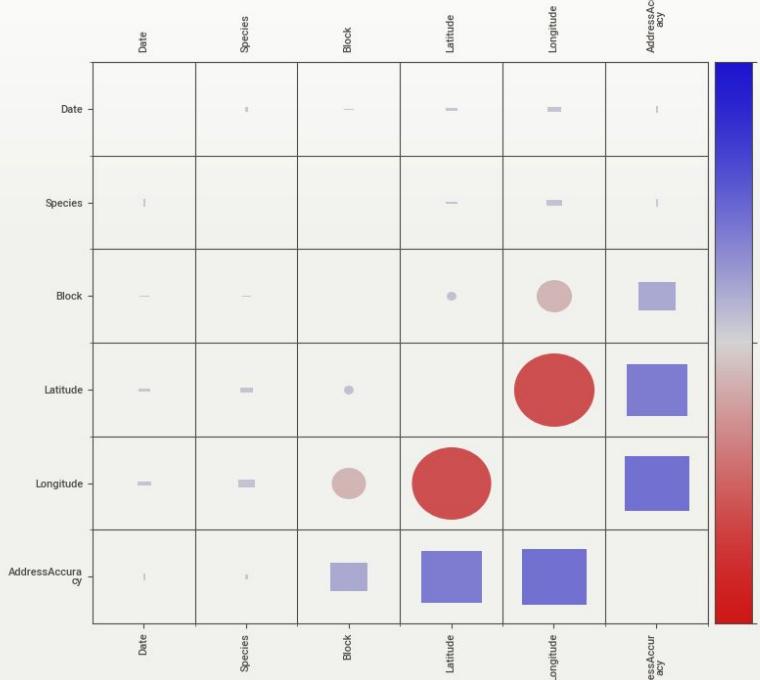
[Only including dataset "Training Data"]
 • Squares are categorical associations (uncertainty coefficient & correlation ratio) from 0 to 1. The uncertainty coefficient is assymmetrical, (i.e. ROW LABEL values indicate how much they PROVIDE INFORMATION to each LABEL at the TOP).

- Circles are the symmetrical numerical correlations (Pearson's) from -1 to 1. The trivial diagonal is intentionally left blank for clarity.



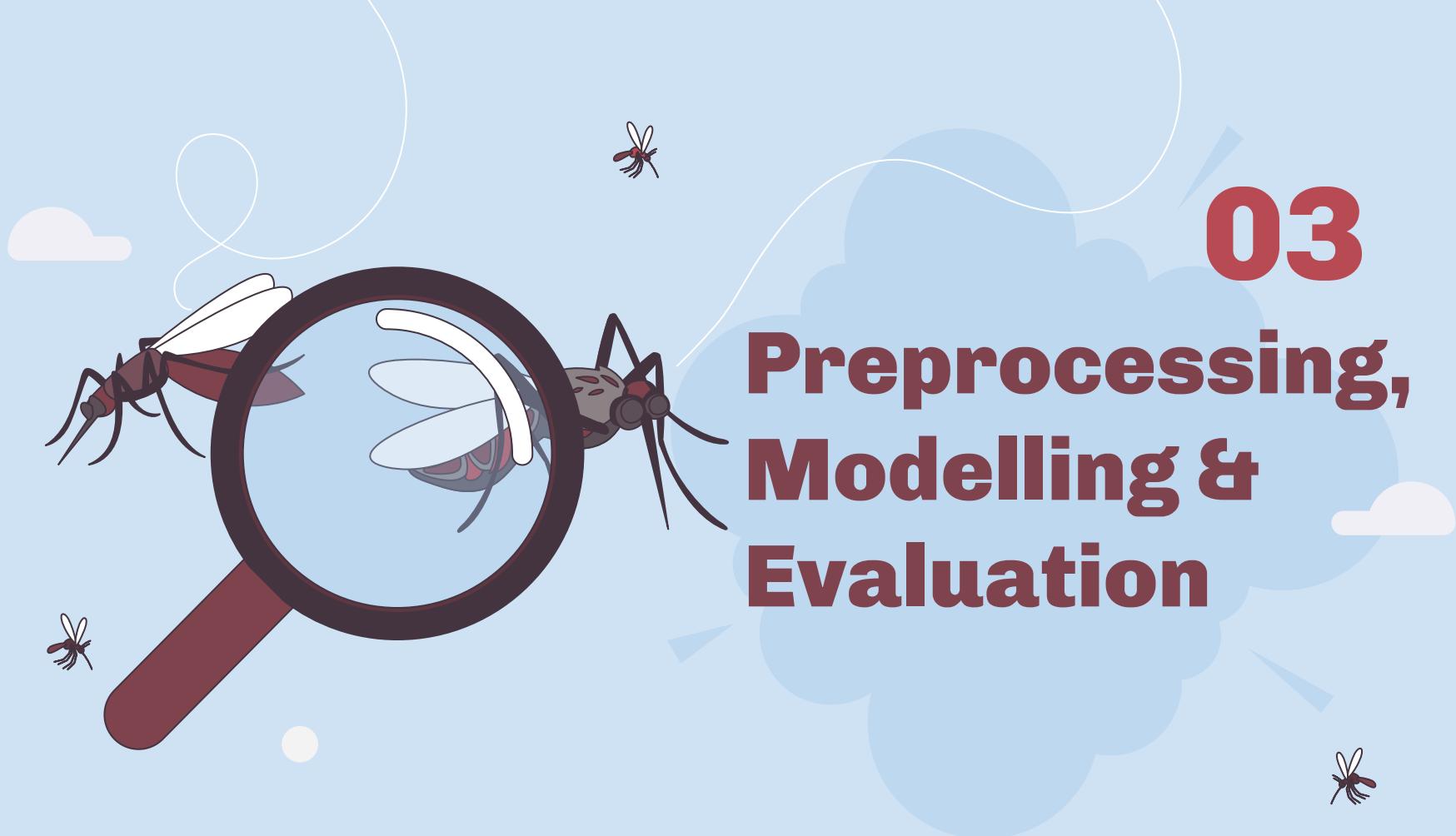
[Only including dataset "Test Data"]
 • Squares are categorical associations (uncertainty coefficient & correlation ratio) from 0 to 1. The uncertainty coefficient is assymmetrical, (i.e. ROW LABEL values indicate how much they PROVIDE INFORMATION to each LABEL at the TOP).

- Circles are the symmetrical numerical correlations (Pearson's) from -1 to 1. The trivial diagonal is intentionally left blank for clarity.



03

Preprocessing, Modelling & Evaluation



Preprocessing

1. **One-hot Encoded categorical features. (ie: Traps, Species)**
2. **Polynomial Features degree = 2**

To uncover potential new relationships between the features and the target.



Feature Engineering



**Longitude
&
Latitude**

Combined



**Sunrise
&
Sunset**

Calculate daylight hours



**Relative
humidity**

Calculated



Feature Engineering

'dew_point_depression'

Average temp. - dew point temp.

'wet_bulb_depression'

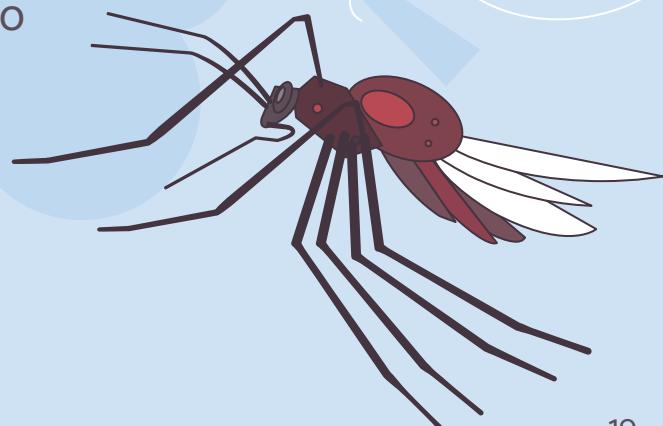
Dry-bulb temp - wet-bulb temp.

Dealing with Imbalanced Data

The train dataset is highly unbalanced, with WnvPositive at 5%.

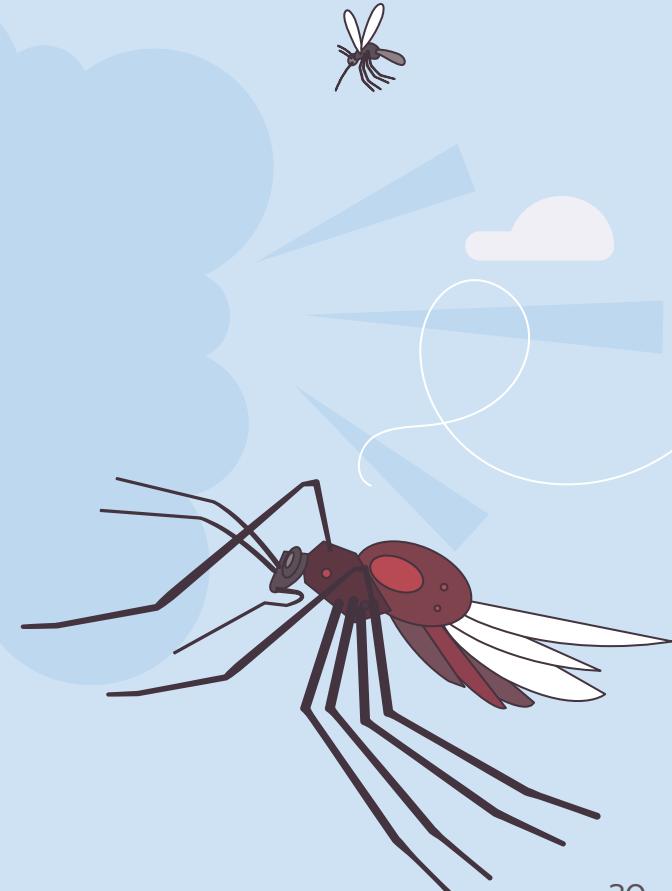
Solutions:

SMOTE (Synthetic Minority Over-sampling Technique) to help with over-sampling.



Models

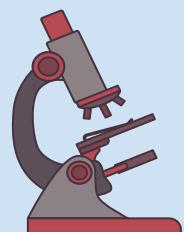
- 1. Logistic Regression**
- 2. GradientBoosting**
- 3. RandomForest**
- 4. AdaBoost**
- 5. ExtraTrees**
- 6. Support Vector Classifier**



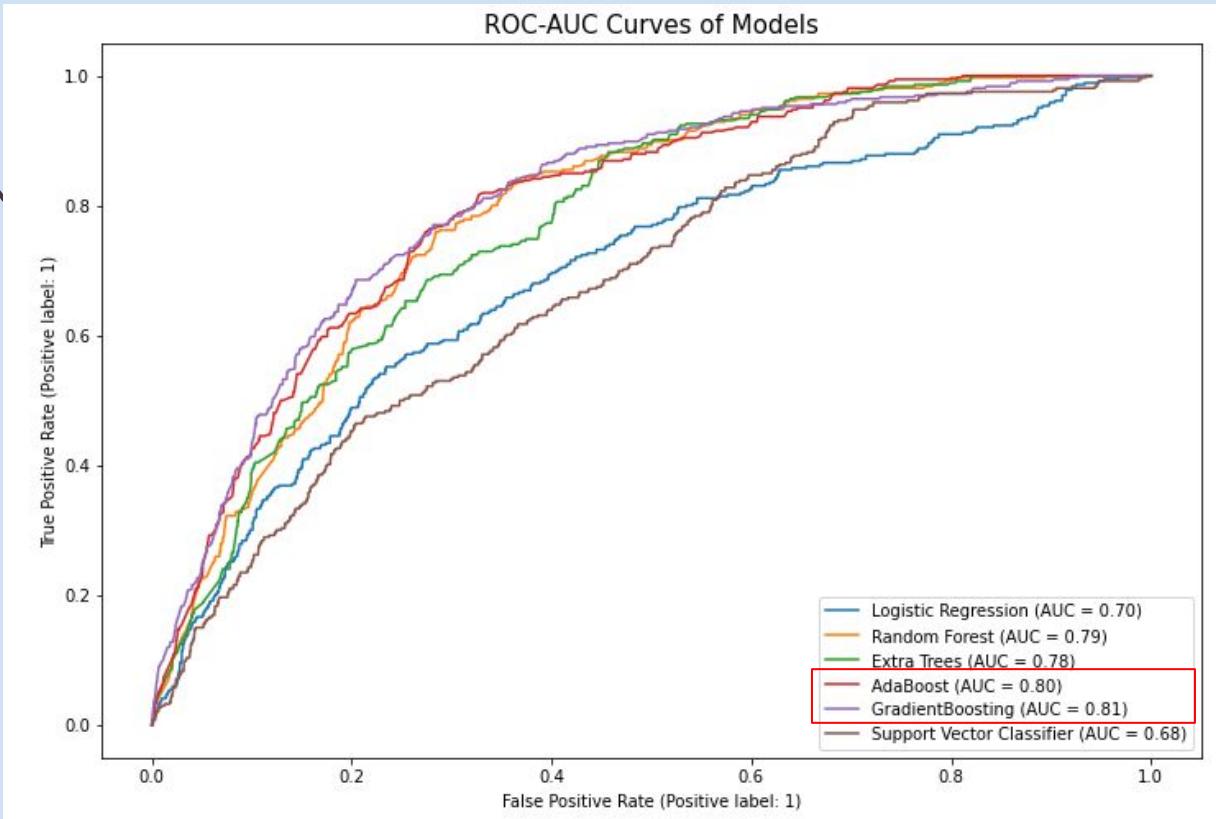
Results ranked by F1-Score



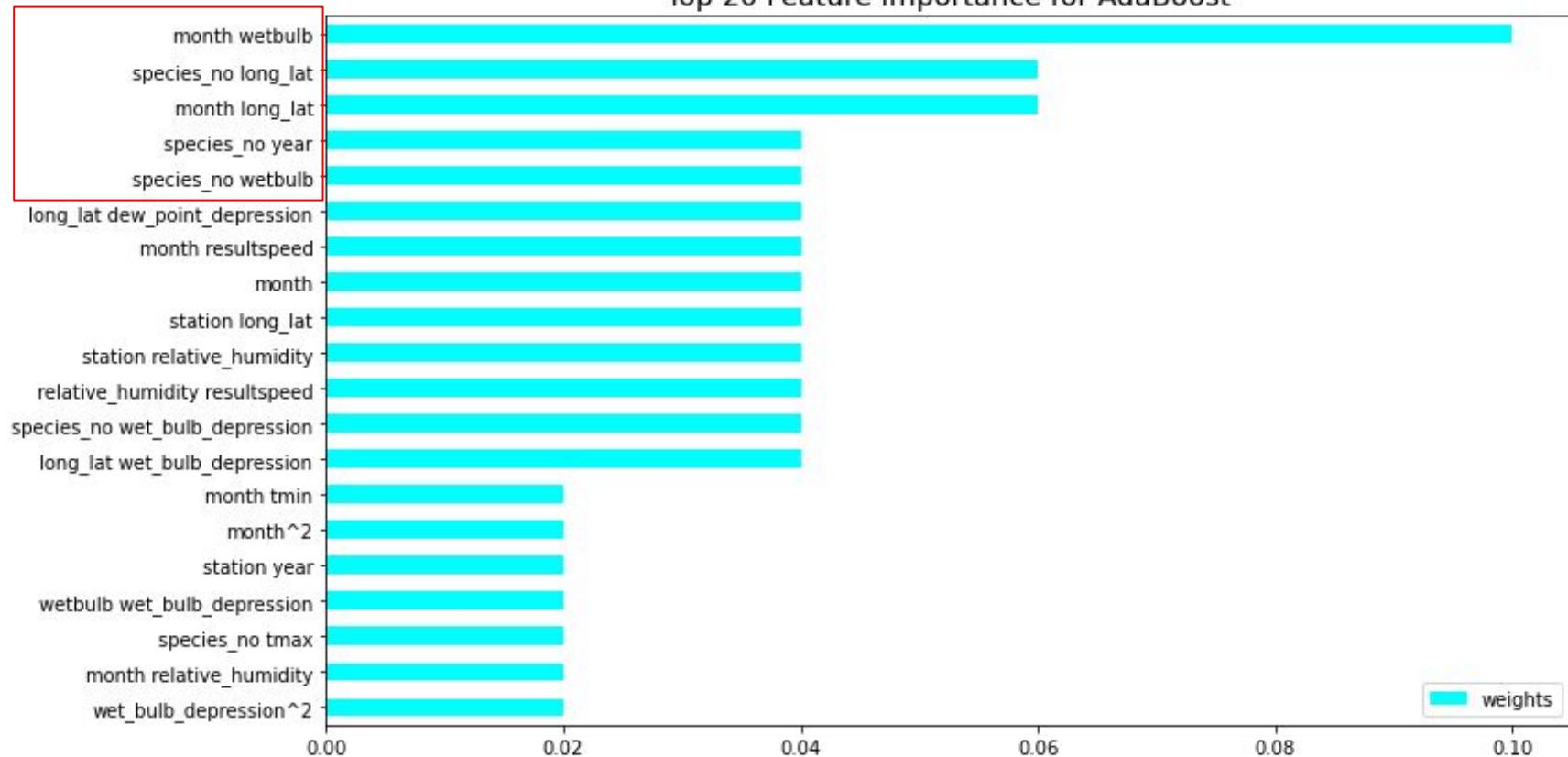
	Model Type	Train AUC	Test AUC	F1-Score	Recall	Precision	Train Accuracy	Test Accuracy
3	AdaBoost	0.782796	0.658550	0.587166	0.658550	0.570800	0.926434	0.860914
4	GradientBoosting	0.794487	0.674291	0.585625	0.674291	0.570438	0.914900	0.849263
1	Random Forest	0.811714	0.711152	0.557298	0.711152	0.561091	0.888092	0.784956
2	Extra Trees	0.812495	0.722152	0.537426	0.722152	0.556724	0.830736	0.744838
0	Logistic Regression	0.809299	0.638893	0.512745	0.638893	0.536702	0.830736	0.740855
5	Support Vector Classifier	0.731718	0.623479	0.460699	0.623479	0.526886	0.774626	0.636136



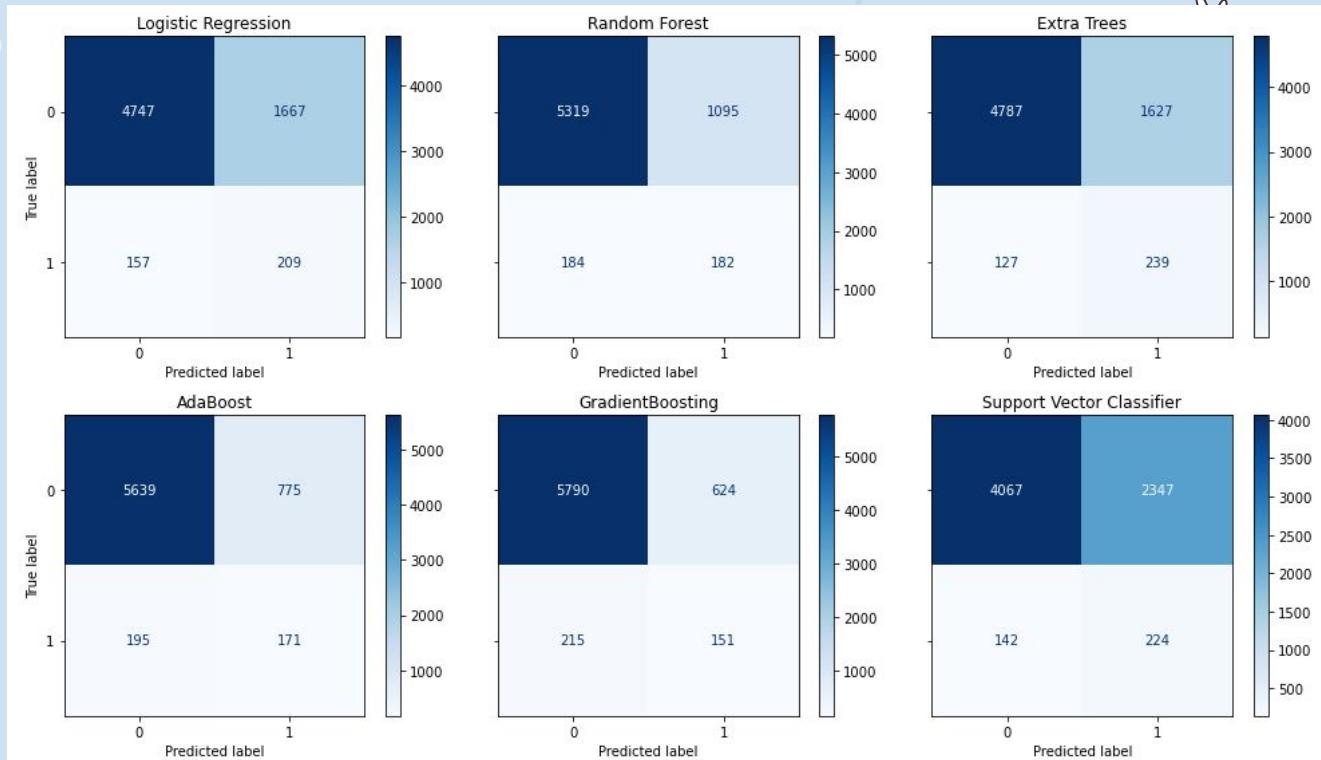
ROC-AUC Curves



Top 20 Feature Importance for AdaBoost



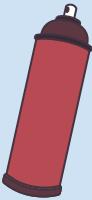
Confusion Matrices



A cartoon illustration of a mosquito being injected by a syringe. The mosquito is black with white wings and is positioned above the needle. The syringe has a red liquid with small white dots, likely representing blood. The background consists of light blue clouds and a few small white dots.

04

Cost Benefit Analysis



Spray costs



The Chicago Department of Public Health (CDPH) is continuing its efforts to combat against the West Nile virus.



The insecticide used by the CDPH is the Zenivex E4, approved for use by the U.S. Environmental Protection Agency and widely applied to control mosquitoes.

Further research¹ showed that the CDPH is spraying approximate 1120 acres per exercise. Assuming at a cost of \$87 per gallon, and with 1.5 fluid ounces per acre used during each exercise, we worked out to approximate **\$1141.88 per exercise.**

1. https://www.chicago.gov/content/dam/city/depts/cdph/CDPH/Healthy%20Chicago/SprayZone_T220_08252021.pdf





Hospitalisation costs

 Based on the report from the U.S. National Library of Medicine National Institute of Health¹, the cost to treat an infected patient diagnosed with West Nile neuroinvasive disease amounts to **~\$33 000 per case**.

For milder cases where the patient only suffers from West Nile Fever, the treatment cost is **~\$300**.

1. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3322011/>



Spray vs medical costs



Based on the number of WNV cases predicted by our model:

	Spray Cost (1 Exercise per month)	Medical Cost (West Nile Fever)	Which is higher?
Year 2008	\$ 13 702.50	\$ 371 460	Medical Cost by \$ 357 757.50
Year 2010	\$ 13 702.50	\$ 578 632	Medical Cost by \$ 564 929.50
Year 2012	\$ 13 702.50	\$ 1 336 954	Medical Cost by \$ 1 323 251.50
Year 2014	\$ 13 702.50	\$ 491 958	Medical Cost by \$ 478 255.50



Spray vs medical costs



Based on the number of WNV cases predicted by our model:

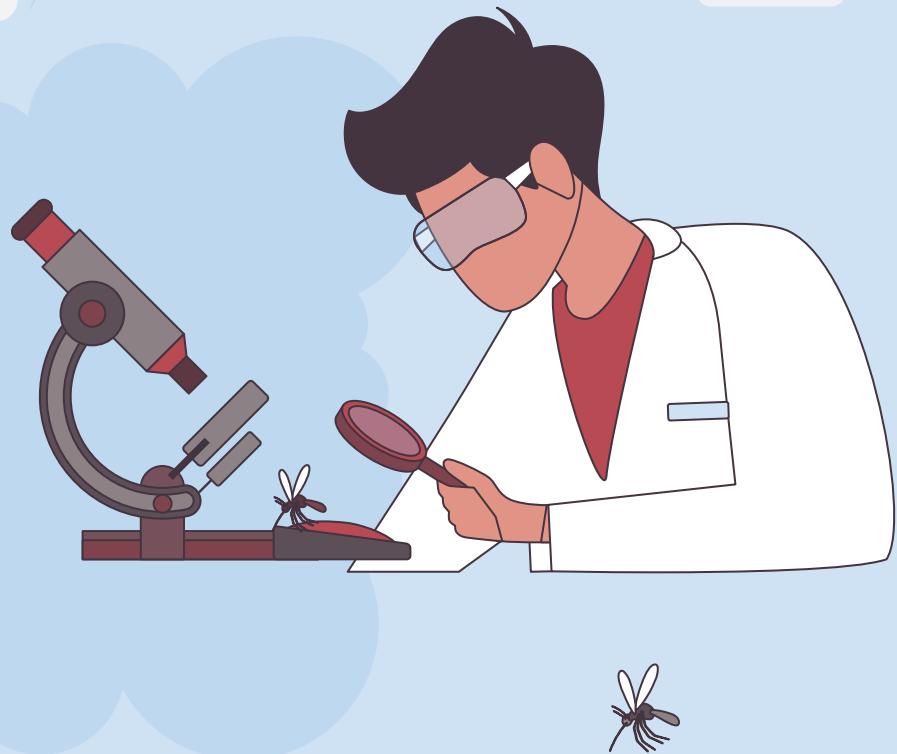
	Spray Cost <small>(33 Exercise per month)</small>	Medical Cost <small>(West Nile Fever)</small>	Which is higher?
Year 2008	\$ 452 182.50	\$ 371 460	Spray Cost by \$ 80 722.50
Year 2010	\$ 452 182.50	\$ 578 632	Medical Cost by \$ 126 449.50
Year 2012	\$ 452 182.50	\$ 1 336 954	Medical Cost by \$ 884 771.50
Year 2014	\$ 452 182.50	\$ 491 958	Medical Cost by \$ 39 775.50

Spray them!!

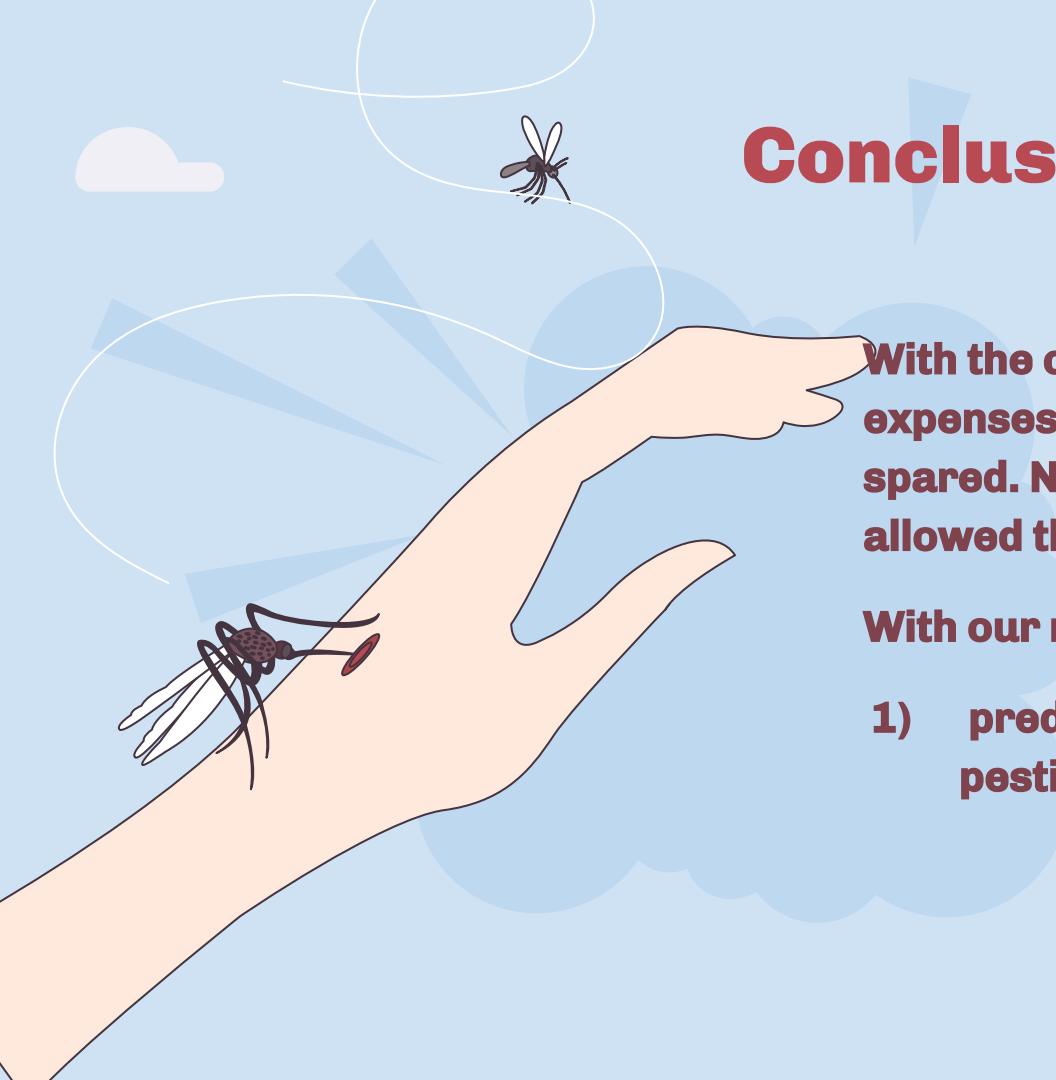


05

Conclusion & Recommendation



Conclusion

A stylized illustration of a human arm and hand. A mosquito is shown on the forearm, and another mosquito is flying above the hand. The background features abstract blue shapes and white clouds.

With the cost-benefit analysis, it is clear that expenses incurred for spraying cannot be spared. Neither can there be an outbreak allowed that will drain the city's resources.

With our model, we can a win-win situation by:

- 1) predicting where and when to spray pesticides,**

\$1.78 mil

Cost per year to spray all traps in entire Chicago

\$400 000

Cost per year to spray 25% of traps with highest WNV virus probability

July, Aug & Sep

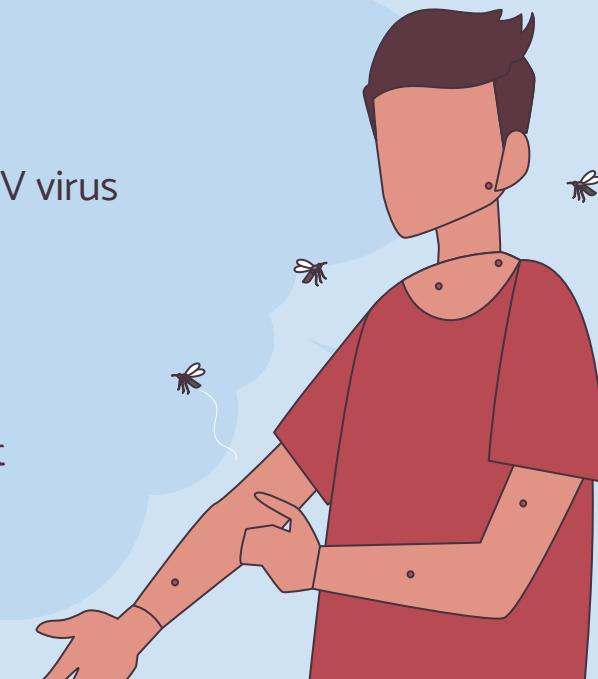
Months to spray when mosquito count is the highest

Where?

- Areas with highest WNV

When?

- Rotate every month so each area gets sprayed 3 times a year

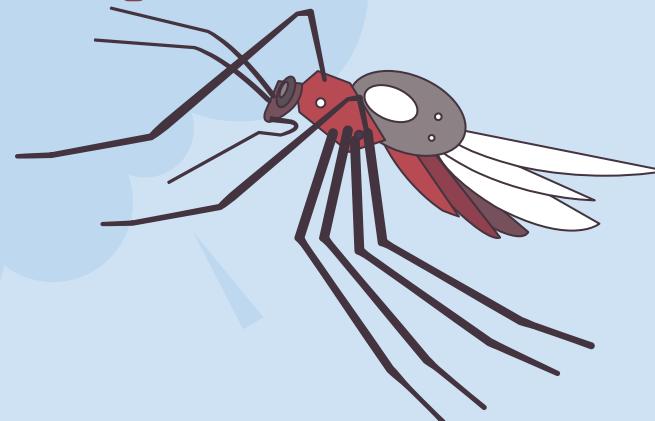


Conclusion



- 2) launch education campaigns at hotspot neighbourhoods to
nip larvae growth in the bud, and

- 3) broadcast reminders to self-protect during months where
mosquito count is expected to peak.





Refinements:



We can further improve by:

- 1) getting more data points to better the accuracy of the prediction of number of mosquitos,**
 - 2) carry out more feature engineering to improve our model score,**
 - 3) gather more data on the spraying efforts to examine its effectiveness regularly, and**
 - 4) collecting more accurate cost calculations that take into account logistics/ manhour costs for the spraying of pesticides.**
- 



Thanks

Do you have any questions?

addyouremail@freepik.com | +91 620 421 838
yourcompany.com

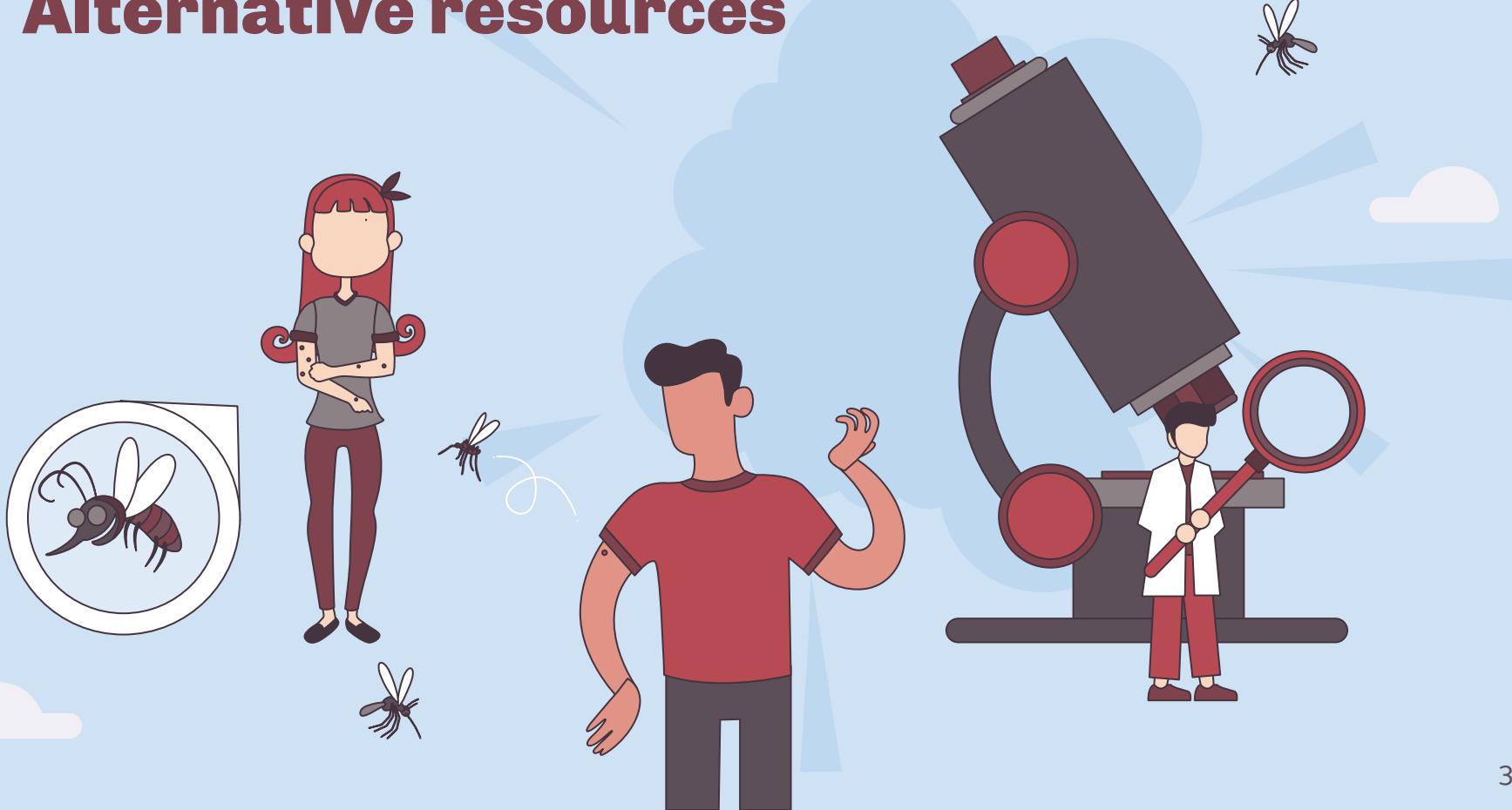


CREDITS: This presentation template was created by Slidesgo, including icons by Flaticon, and infographics & images by Freepik

Please keep this slide for the attribution



Alternative resources



Alternative resources



Vectors:

- Mosquito biting person with flat design
- Mosquito biting person
- Flat design biotechnology concept illustrated

Photos:

- Close up microscope glass slide
- Medium shot scientists laboratory
- Close up researcher working with microscope
- Front view science elements with chemicals composition
- Scientists working with microscope
- Defocused female scientist with safety glasses
- Doctor analyzing with microscope



Resources



Vectors:

- Mosquito control background
- Mosquito biting person with flat design
- Hand with mosquito spray
- Hand drawn composition mosquito biting
- Mosquito biting person with flat design
- Mosquito biting person
- Mosquito biting with flat design
- Mosquito biting person with flat

- Coronavirus vaccine phases info
- Hand drawn composition mosquito biting
- Science concept with microscope
- Flat design scientific research
- Hand drawn composition mosquito
- Mosquito biting person with flat
- Mosquito control background
- Mosquito biting
- Flat illustration biotechnology concept
- Flat biotechnology concept illustrated
- Health clinic landing page template



Instructions for use

In order to use this template, you must credit **Slidesgo** by keeping the **Thanks** slide.

You are allowed to:

- Modify this template.
- Use it for both personal and commercial projects.

You are not allowed to:

- Sublicense, sell or rent any of Slidesgo Content (or a modified version of Slidesgo Content).
- Distribute Slidesgo Content unless it has been expressly authorized by Slidesgo.
- Include Slidesgo Content in an online or offline database or file.
- Offer Slidesgo templates (or modified versions of Slidesgo templates) for download.
- Acquire the copyright of Slidesgo Content.

For more information about editing slides, please read our FAQs or visit Slidesgo School:

<https://slidesgo.com/faqs> and <https://slidesgo.com/slidesgo-school>

Instructions for use (premium users)

In order to use this template, you must be a Premium user on [Slidesgo](#).

You are allowed to:

- Modify this template.
- Use it for both personal and commercial purposes.
- Hide or delete the “Thanks” slide and the mention to Slidesgo in the credits.
- Share this template in an editable format with people who are not part of your team.

You are not allowed to:

- Sublicense, sell or rent this Slidesgo Template (or a modified version of this Slidesgo Template).
- Distribute this Slidesgo Template (or a modified version of this Slidesgo Template) or include it in a database or in any other product or service that offers downloadable images, icons or presentations that may be subject to distribution or resale.
- Use any of the elements that are part of this Slidesgo Template in an isolated and separated way from this Template.
- Register any of the elements that are part of this template as a trademark or logo, or register it as a work in an intellectual property registry or similar.

For more information about editing slides, please read our FAQs or visit Slidesgo School:

<https://slidesgo.com/faqs> and <https://slidesgo.com/slidesgo-school>

Fonts & colors used

This presentation has been made using the following fonts:

Chivo

(<https://fonts.google.com/specimen/Chivo>)

Catamaran

(<https://fonts.google.com/specimen/Catamaran>)

#443440

#5c4f58

#5d3742

#8c8184

#76515b

#7f434e

#b84a54

#cfe2f3

#ffffff

Storyset

Create your Story with our illustrated concepts. Choose the style you like the most, edit its colors, pick the background and layers you want to show and bring them to life with the animator panel! It will boost your presentation. Check out [How it Works](#).



Pana



Amico



Bro



Rafiki

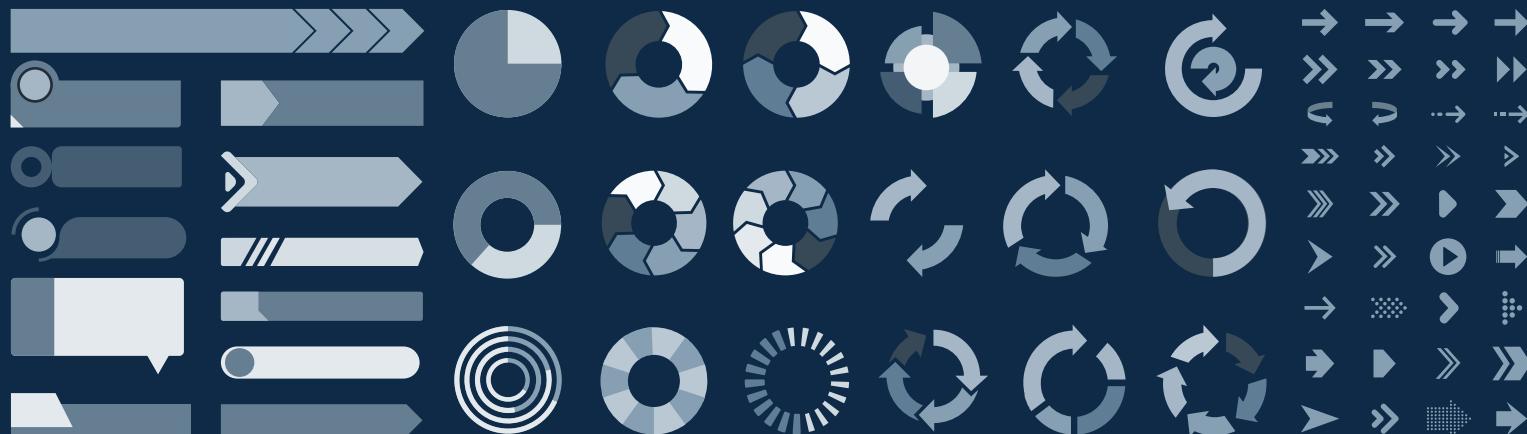


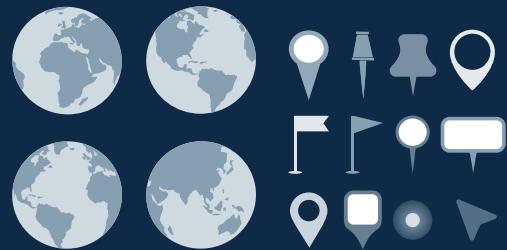
Cuate

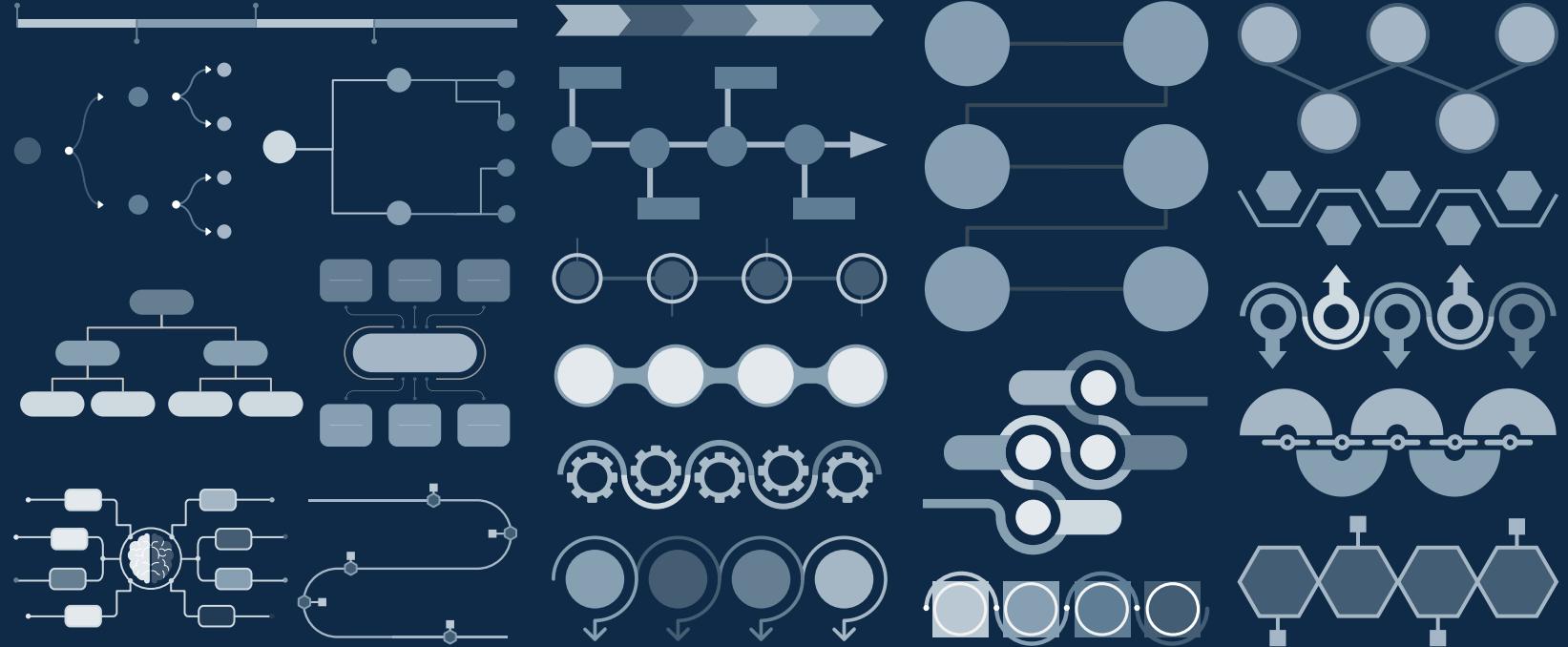
Use our editable graphic resources...

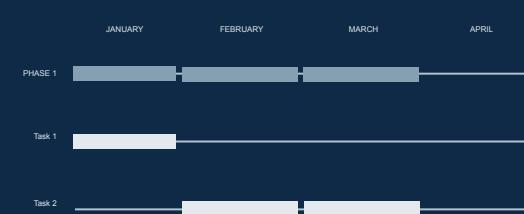
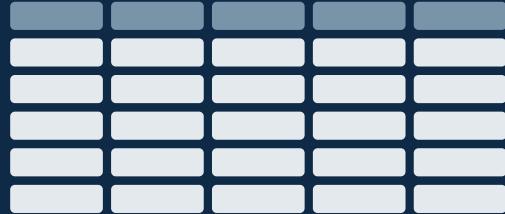
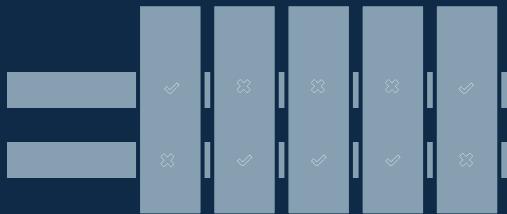
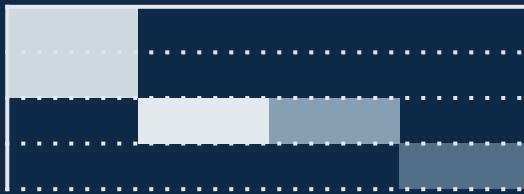
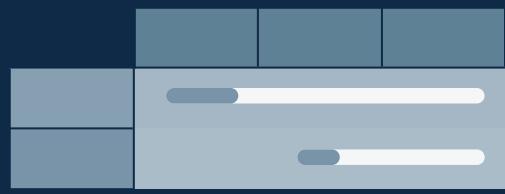
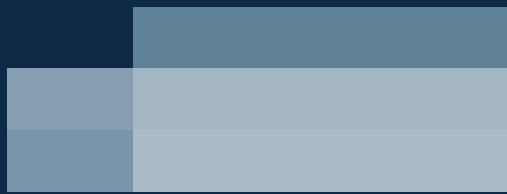
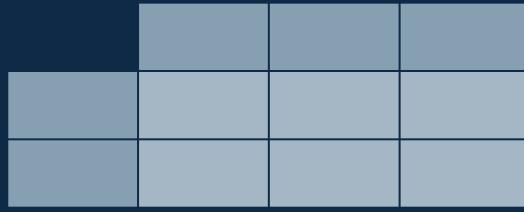
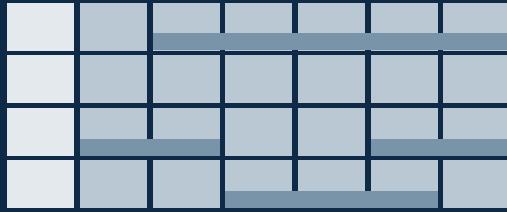
You can easily resize these resources without losing quality. To change the color, just ungroup the resource and click on the object you want to change. Then, click on the paint bucket and select the color you want.

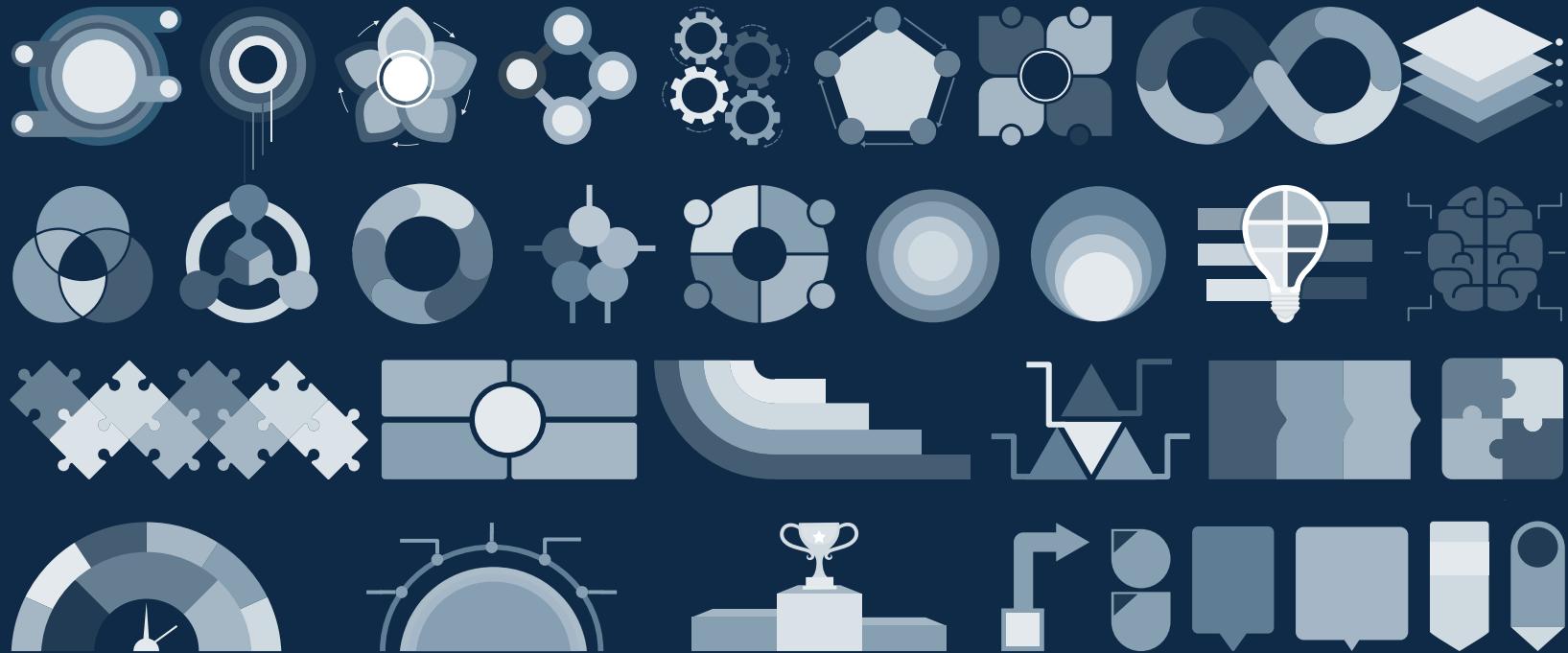
Group the resource again when you're done. You can also look for more infographics on Slidesgo.

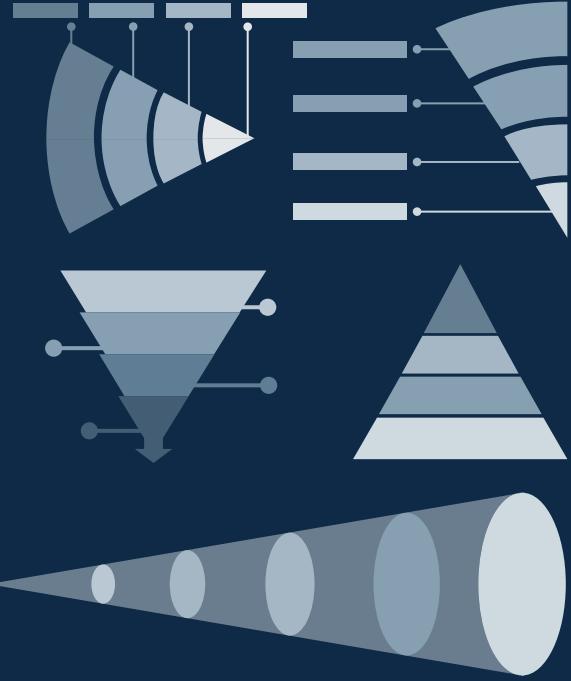
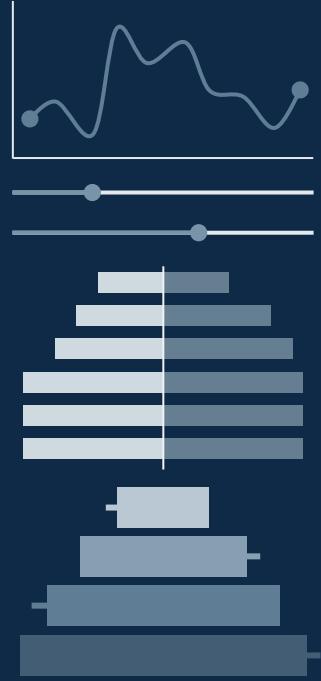
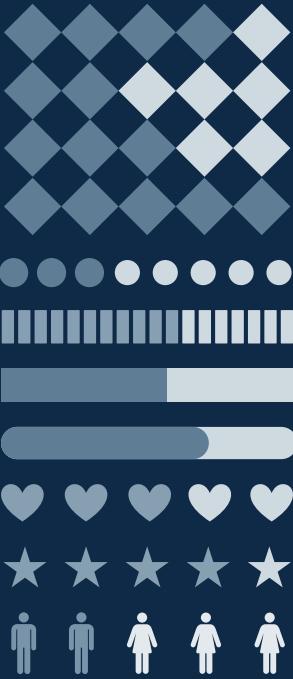
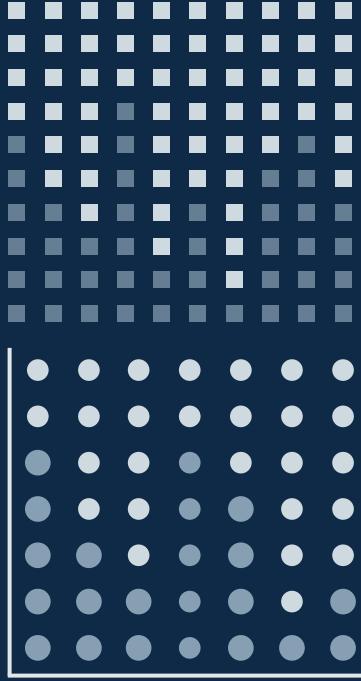












...and our sets of editable icons

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





Contents of this template

Here's what you'll find in this **Slidesgo** template:

1. A slide structure based on a multi-purpose presentation, which you can easily adapt to your needs. For more info on how to edit the template, please visit **Slidesgo School** or read our **FAQs**.
2. An assortment of graphic resources that are suitable for use in the presentation can be found in the **alternative resources** slides.
3. A **thanks** slide, which you must keep so that proper credits for our design are given.
4. A **resources** slide, where you'll find links to all the elements used in the template.
5. **Instructions for use**.
6. Final slides with:
 - The **fonts and colors** used in the template.
 - A selection of **illustrations**. You can also customize and animate them as you wish with the online editor. Visit **Storyset** to find more.
 - More **infographic resources**, whose size and color can be edited.
 - Sets of **customizable icons** of the following themes: general, business, avatar, creative process, education, help & support, medical, nature, performing arts, SEO & marketing, and teamwork.

You can delete this slide when you're done editing the presentation.

Whoa!

This can be the part of the presentation where you can introduce yourself, write your email...





“This is a quote, words full of wisdom that someone important said and can make the reader get inspired”

—**Someone Famous**

Awesome Words





A picture always reinforces the concept

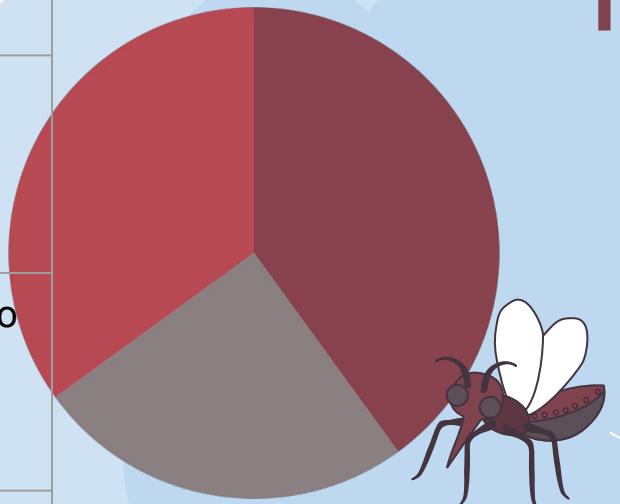
Images reveal large amounts of data, so remember:
use an image instead of a long text



Data Cleaning

Weather

Columns	Action
'sunrise' 'sunset'	No data from station 2 so impute from station 1
'water1' 'codesum' 'snowfall' 'depth'	0 or missing so drop
'Preciptotal' 'avgspeed' 'sealevel' 'stnpressure'	Impute with median values

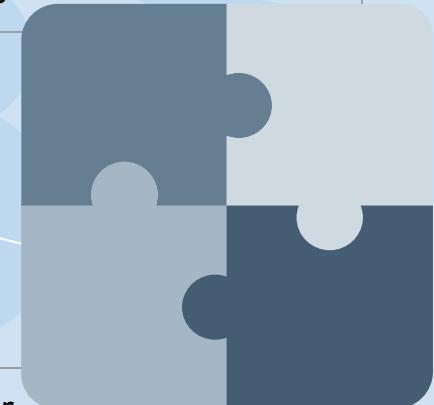


Spray

3 columns of day, month and year

Train-test

No. of mosquitoes that exceed 50 form new row so merge duplicated rows





**A picture is
worth a
thousand
words**



Feature Engineering

Mars

Despite being red, Mars is a cold place

Venus

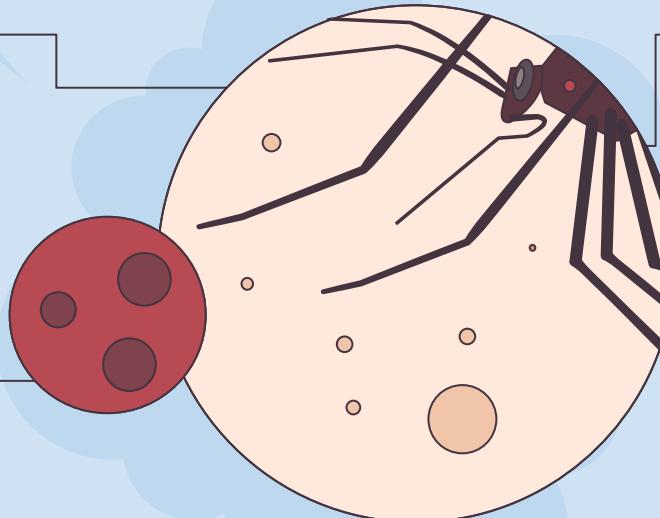
Venus is the second planet from the Sun

Neptune

Neptune is the farthest from the Sun

Jupiter

It's the biggest planet in the Solar System

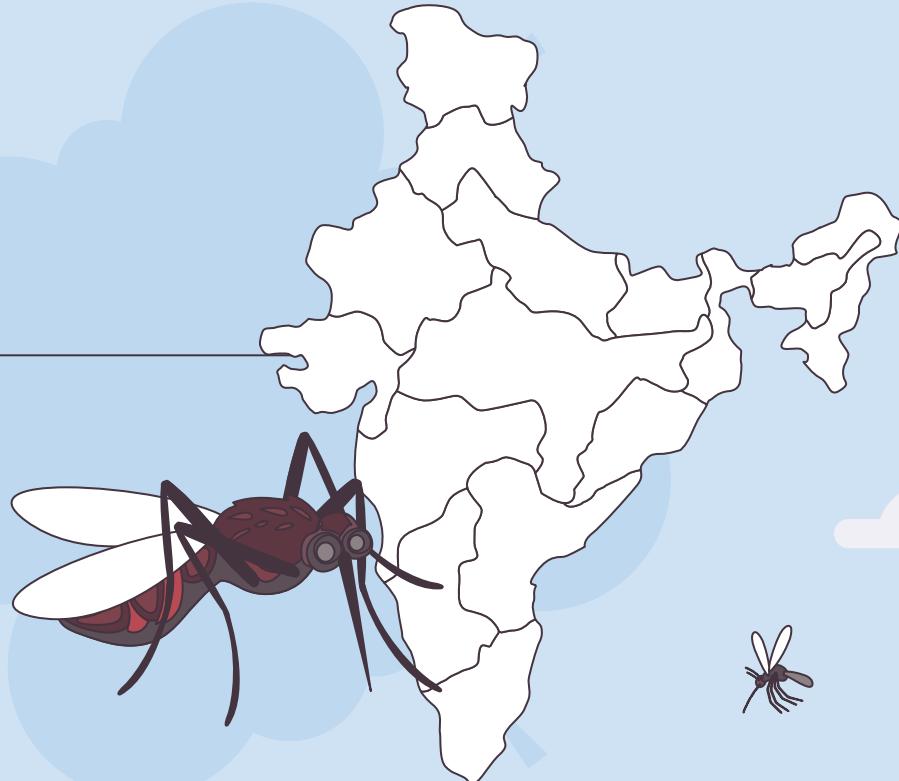


This is a map

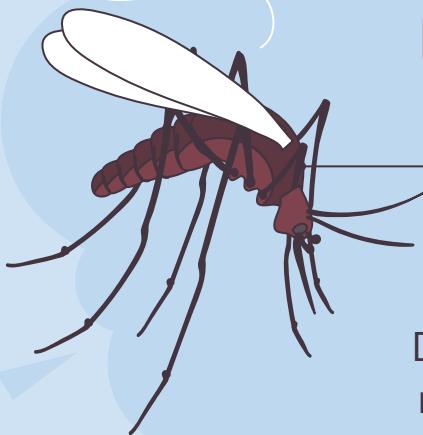


India

It was the country where a disease was first linked to mosquitoes in the 1970s



A timeline always works well



Phase 1

Eggs

Despite being red, Mars is a cold place

Phase 2

Larvae

It has a nice name, but it's terribly hot

Phase 3

Pupae

Mercury is the closest planet to the Sun

Phase 4

Adult

It's the biggest planet in the Solar System



Calendar



August

Mo	Tu	We	Th	Fr	Sa	Su
						1
2	3	4	5	6	7	8
9	10	11	12	13	14	15
16	17	18	19	20	21	22
23	24	25	26	27	28	29
30	31					

20th of August

World Mosquito Day is celebrated, commemorating the discovery of the British physician Sir Ronald Ross in 1897

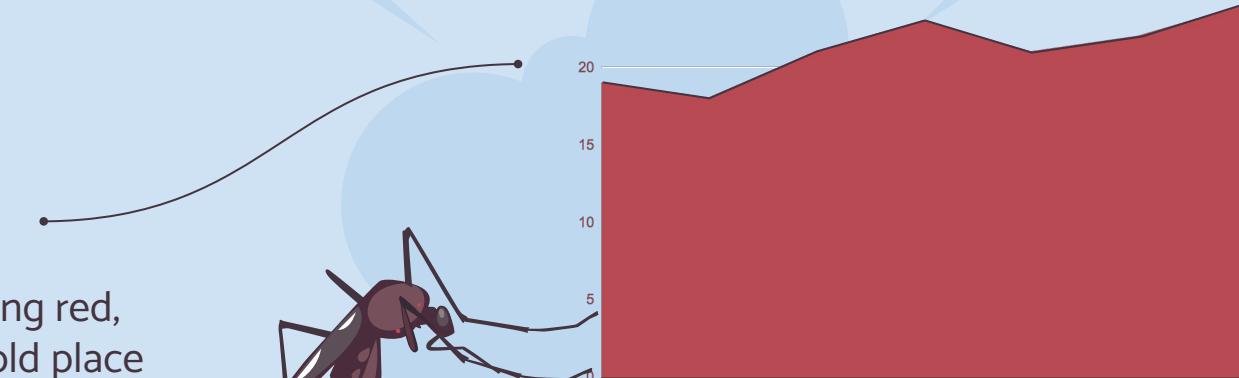


This is another graph



Mars

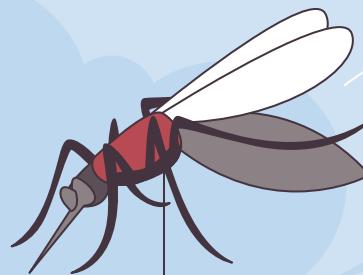
Despite being red,
Mars is a cold place



Follow the link in the graph to modify its data and then paste
the new one here. [For more info, click here](#)



This is a diagram



Mercury

Mercury is the closest
to the Sun

Mars

Despite being red, Mars
is a cold place

Venus

Venus is the second
planet from the Sun



Conclusions



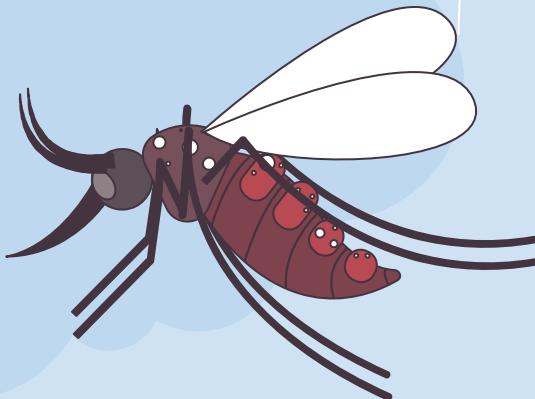
Mercury

Mercury is the closest planet to the Sun



Venus

Venus is the second planet from the Sun



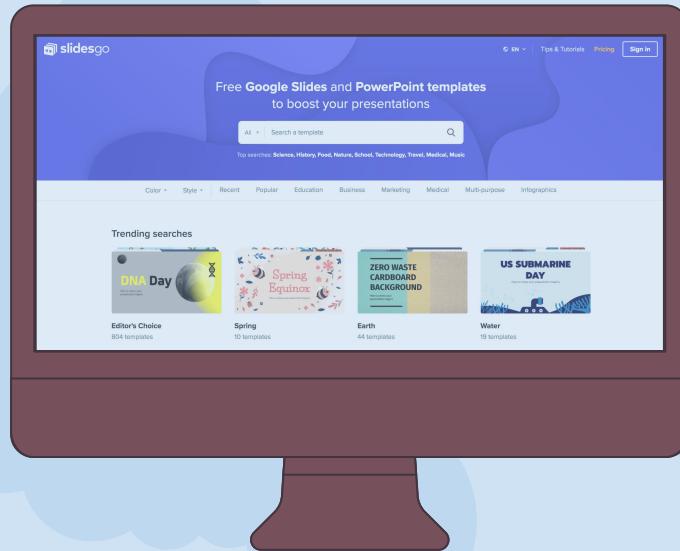
Mars

Despite being red, Mars is a cold place

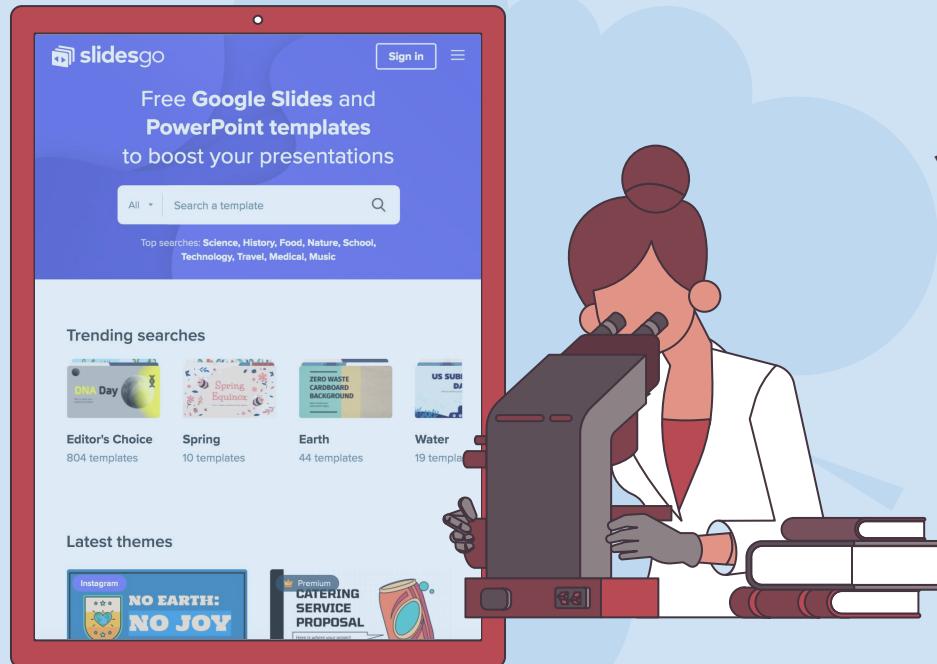


Desktop software

You can replace the image on the screen with your own work. Just delete this one, add yours and center it properly



Tablet app



You can replace the image on the screen with your own work. Just delete this one, add yours and center it properly



Mobile web

You can replace the image on the screen with your own work. Just delete this one, add yours and center it properly

