

# Predicting Ground Level Air Pollution

Anywhere on  
Earth.

David Djambazov, Jerico Johns,  
Srishti Mehra



# The Team



**SRISHTI  
MEHRA**

ML Interests:

- Applied Research
- NLP



**JERICO  
JOHNS**

ML Interests:

- Environmental Advocacy
- Algorithmic Bias

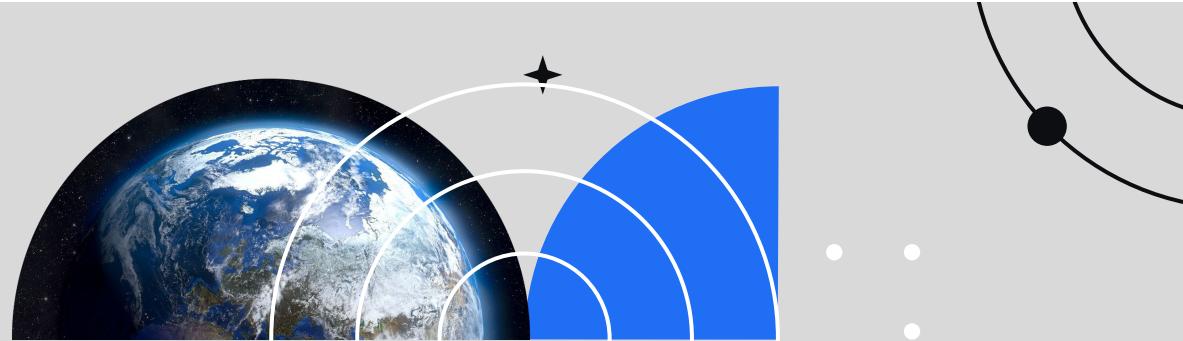


**DAVID  
DJAMBAZOV**

ML Interests:

- Open Source Intelligence
- NLP

# Agenda



## 01 PROBLEM

Problem Description

## 02 SOLUTION

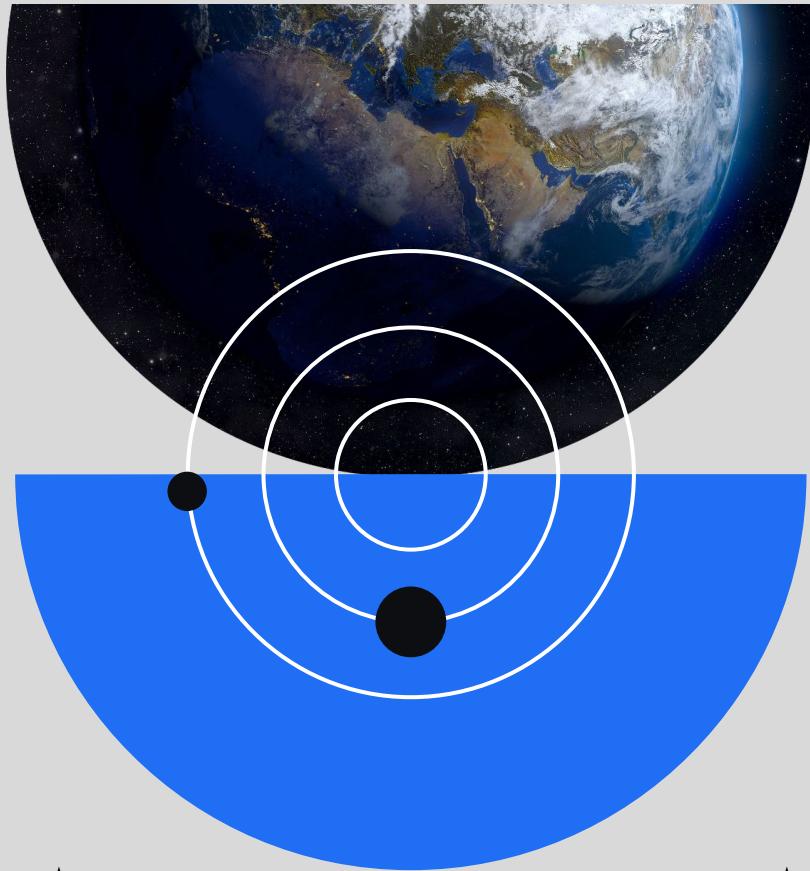
Web App Demo

## 03 HOW IT WORKS

Technical Approach

## 04 Q&A





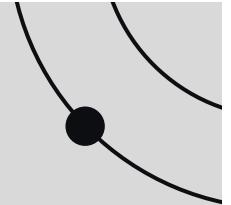
01

# Problem Description



# User Perspective

1



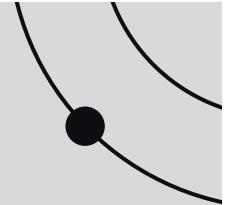
# User Perspective

2

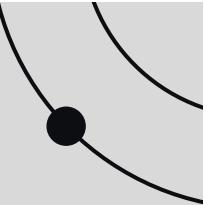


# User Perspective

3



# User Perspective



Los Angeles Times

  Bay Area has the world's worst air quality thanks to this week's Northern California fires



The San Francisco skyline, with the Golden Gate Bridge in the foreground, is obscured by smoke from wildfires on Wednesday. (Eric Risberg / Associated Press)

**SUBSCRIBERS ARE READING >**

MOVIES

FOR SUBSCRIBERS

Here's what really happened after Will Smith slapped Chris Rock at the Oscars

TRAVEL

FOR SUBSCRIBERS

9 magical swimming holes in California to escape the heat

HOUSING & HOMELESSNESS

FOR SUBSCRIBERS

Bidding wars make buying a house in L.A. a nightmare. These buyers found a way around it

LIFESTYLE

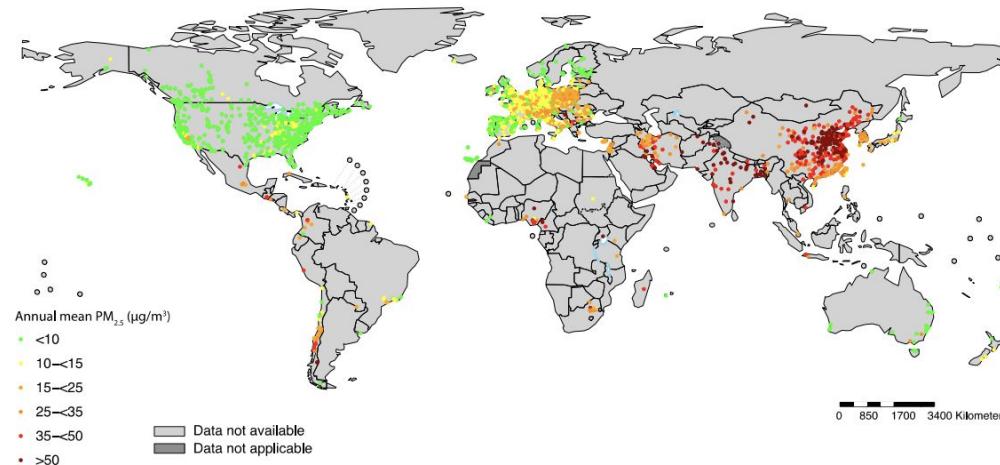
FOR SUBSCRIBERS

They were spending all their income on rent. A garage turned ADU saved them

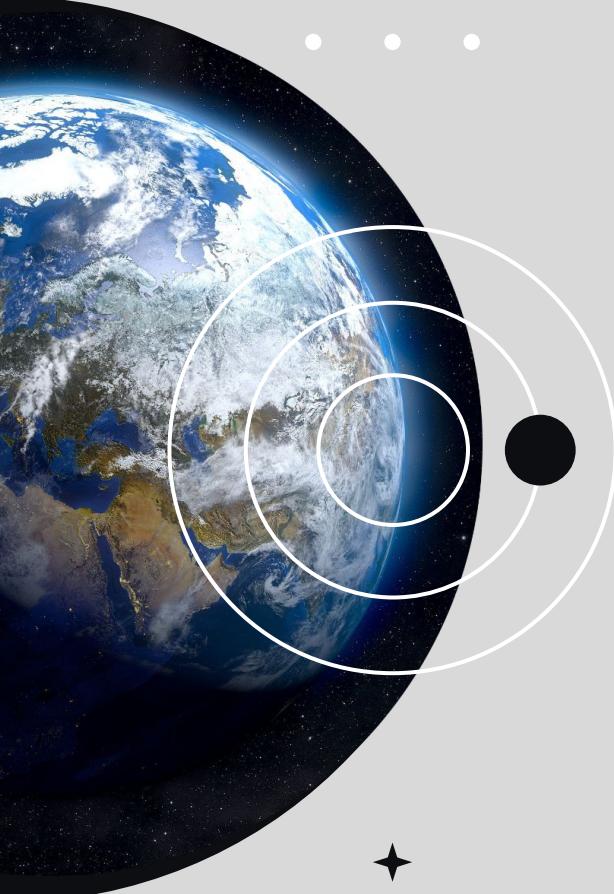


# Distribution of PM<sub>2.5</sub> Monitoring Stations is concentrated in developed, urban areas.

Figure 1: Location of the monitoring stations and PM<sub>2.5</sub> concentration in more than 4000 human settlements, 2010-2016



PM<sub>2.5</sub>: Fine particulate matter of a diameter of 2.5 microns or less.



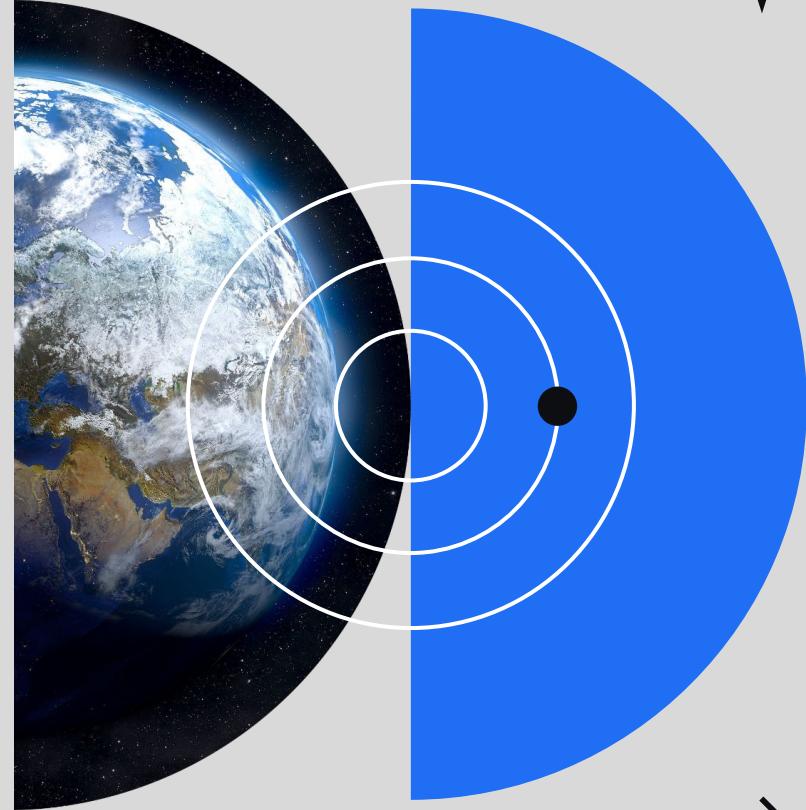
7 million  
deaths a  
year.

*Air pollution is the leading environmental cause of human deaths ([WHO](#)).*

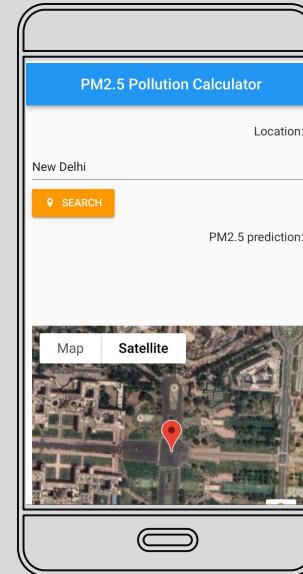
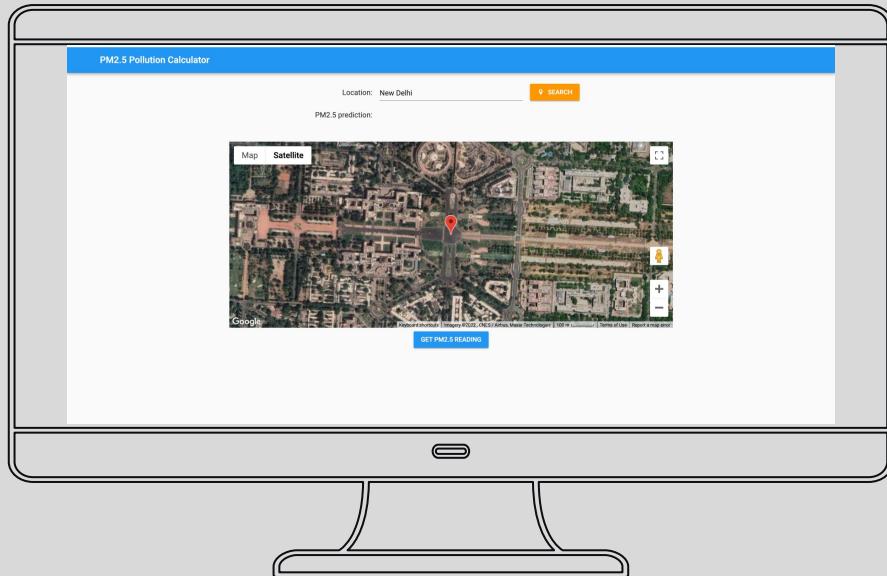


# 02

## Solution



# Web App Demo



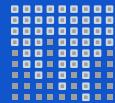
# How We Got Here

5 TB Data Scraped



50+ Research  
Papers

5 Imputation  
Strategies



1,000 Feature  
Combinations

3 Cross-Validation  
Strategies

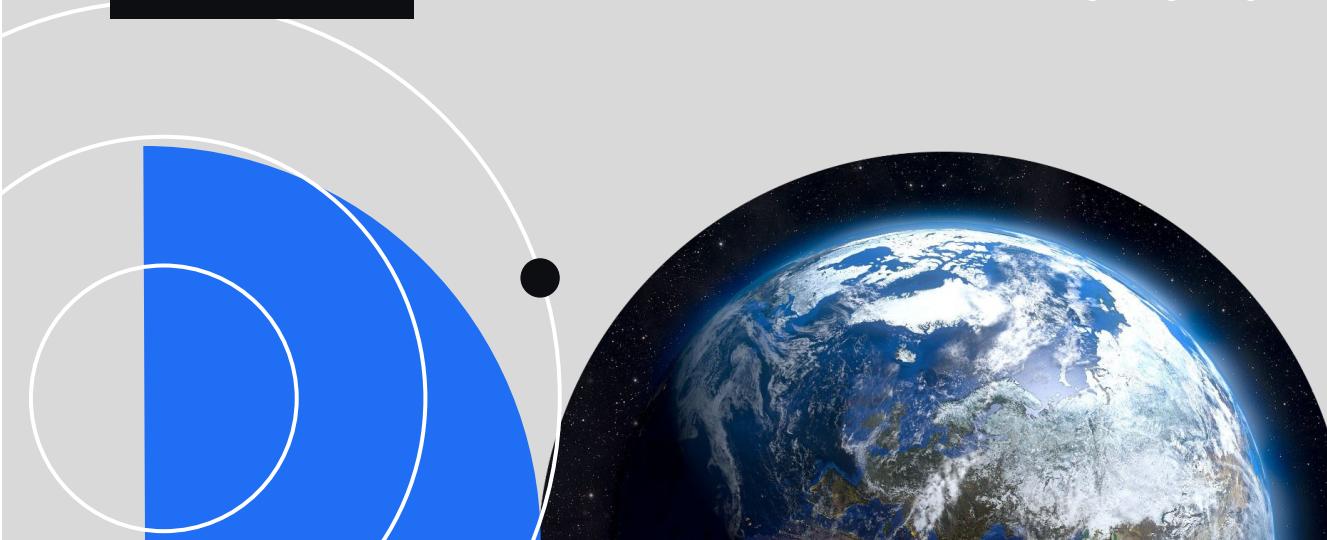


7 Model  
Architectures

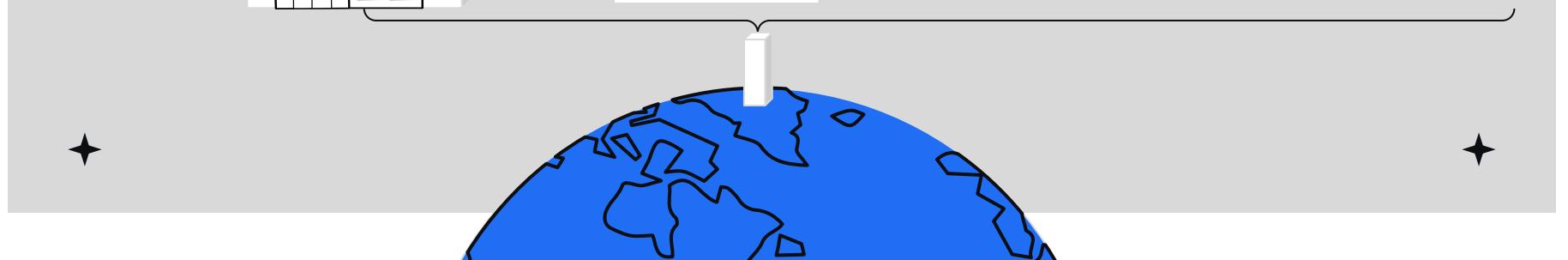
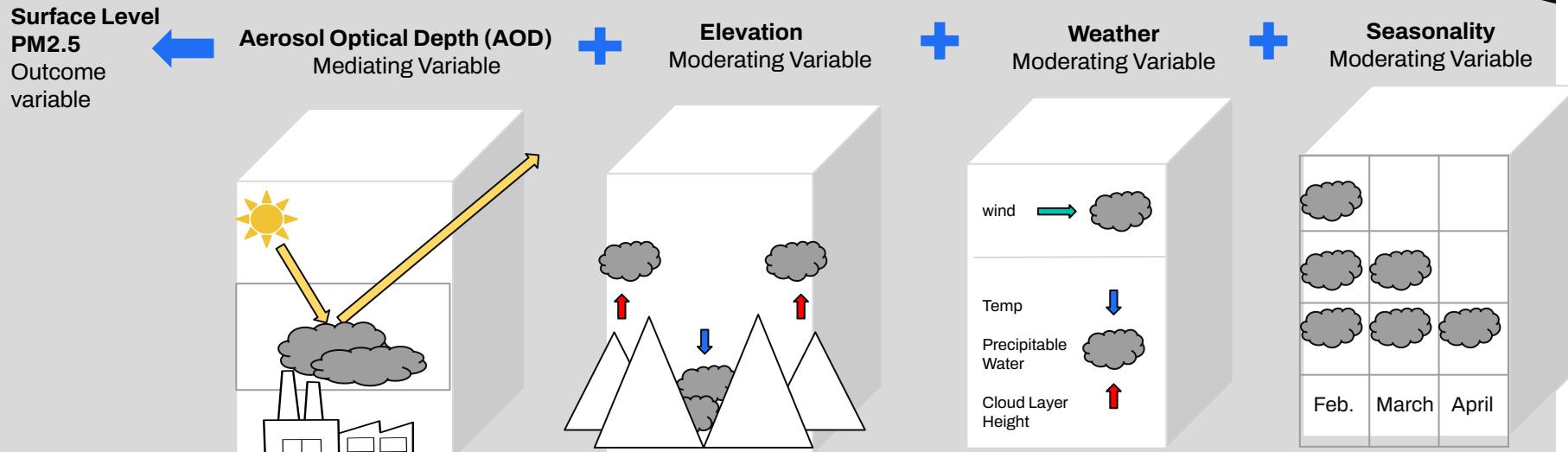


03

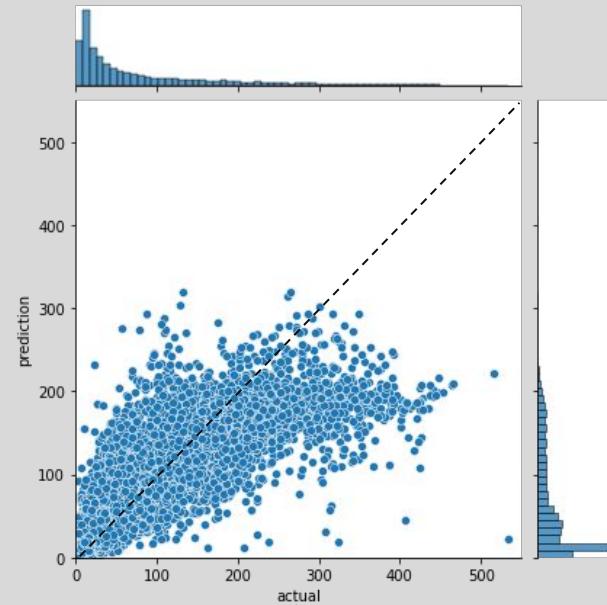
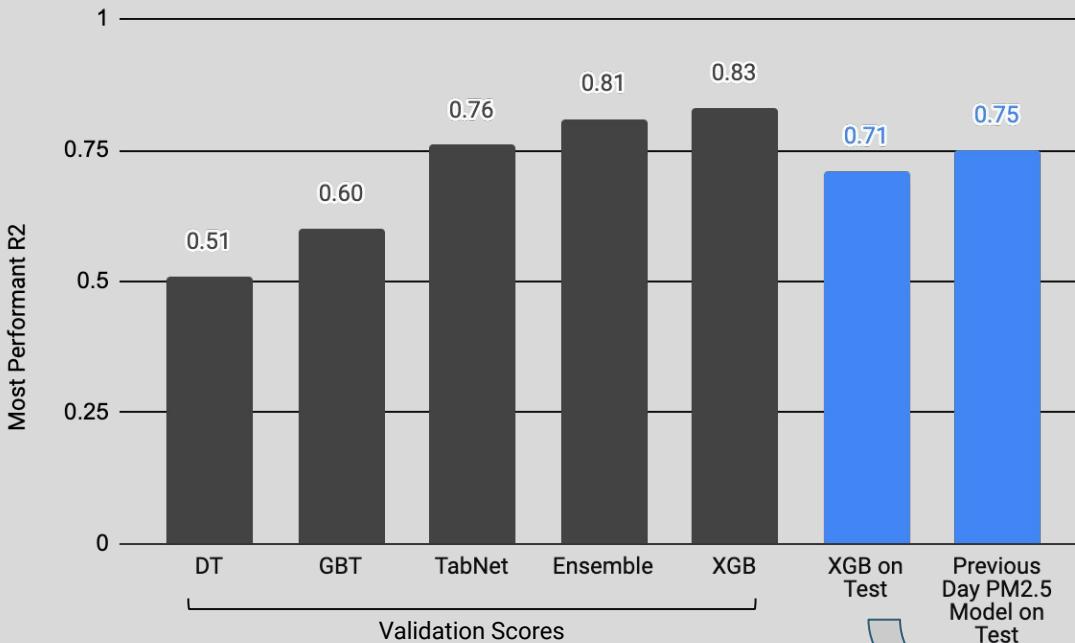
# How It Works



# A Scalable Approach



# Model Performance by R<sup>2</sup> Score

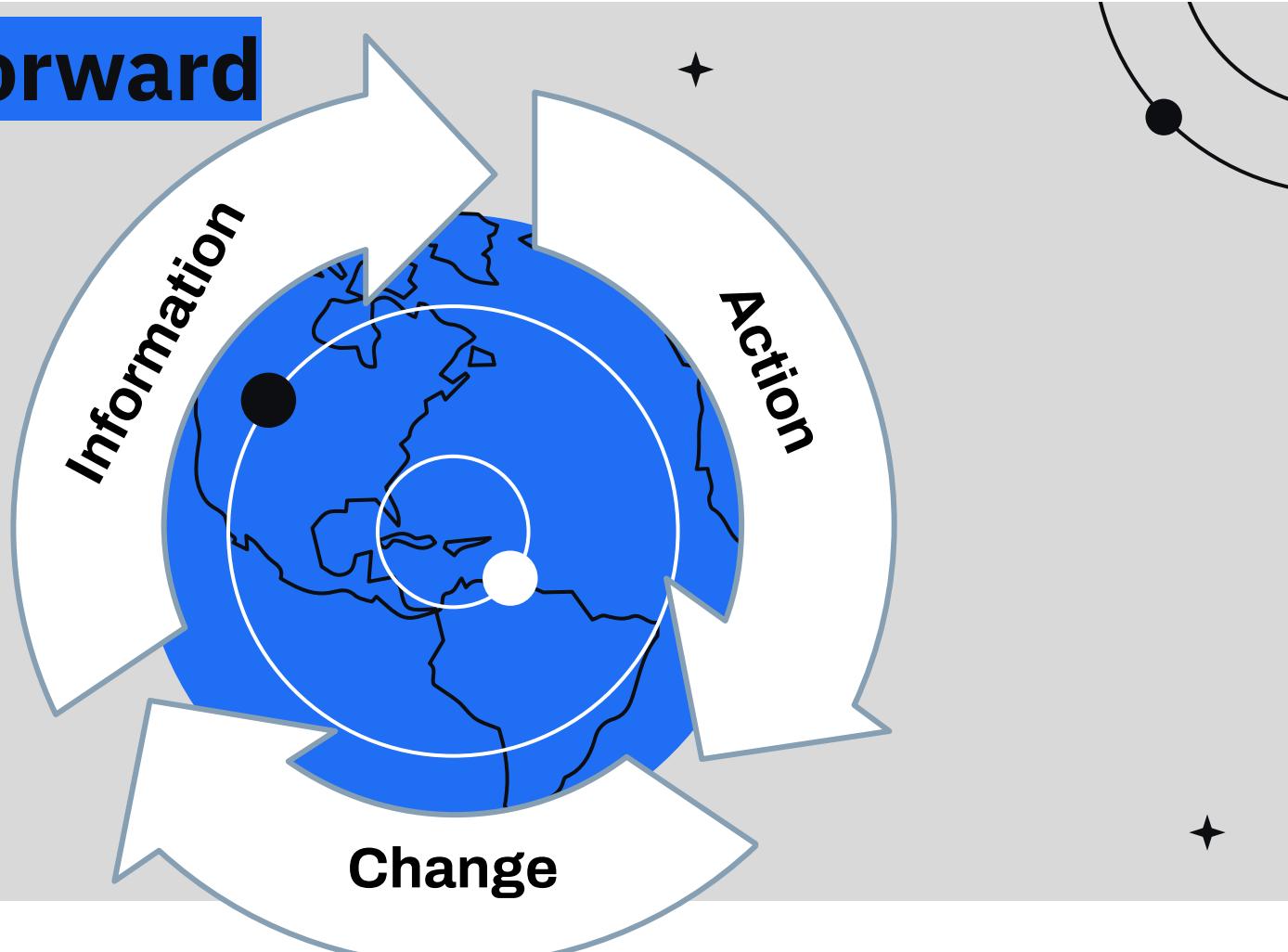
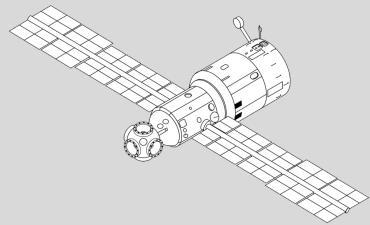


# Translating Performance



AQI Category	Health Impact	PM2.5 (ug/m3)	Prediction Mean (Prediction Mean - Actual Mean)	% of Observations
Good	Minimal	0-30	19.4 (+5.0)	44%
Satisfactory	Minor Breathing discomfort to sensitive people.	31-60	56.9 (+14.0)	17%
Moderately Polluted	Breathing discomfort to asthma patients, elderly, and children.	61-90	83.8 (+10.2)	9%
Poor	Breathing discomfort to all.	91-120	104.7 (+4.6)	6%
Very Poor	Respiratory illness on prolonged exposure.	121-250	147.5 (-28.6)	17%
Severe	Health impact even on light physical work. Serious impact on people with heart/lung disease.	250+	184.5 (-127.0)	6%
Total			68.8 (-6.3)	10,225

# A Path Forward



# Q&A



# References

## Helpful Papers:

- [Estimating Regional Spatial and Temporal Variability of PM2.5 Concentrations Using Satellite Data, Meteorology, and Land Use Information](#)
- [Estimating ground-level PM<sub>2.5</sub> concentrations in the Southeastern United States using MAIAC AOD retrievals and a two-stage model](#)
- [What Can Affect AOD–PM<sub>2.5</sub> Association?](#)
- [Comparison of GOES and MODIS Aerosol Optical Depth \(AOD\) to Aerosol Robotic Network \(AERONET\) AOD and IMPROVE PM<sub>2.5</sub> Mass at Bondville, Illinois](#)
- [Intercomparison between satellite-derived aerosol optical thickness and PM<sub>2.5</sub> mass: Implications for air quality studies](#)
- [A Review on Predicting Ground PM<sub>2.5</sub> Concentration Using Satellite Aerosol Optical Depth](#)
- [A Robust Deep Learning Approach for Spatiotemporal Estimation of Satellite AOD and PM<sub>2.5</sub>](#)
- [Prediction algorithm of PM<sub>2.5</sub> mass concentration based on adaptive BP neural network](#)
- [Satellite-Based Spatiotemporal Trends in PM<sub>2.5</sub> Concentrations: China, 2004–2013](#)
- [PM<sub>2.5</sub>-GNN: A Domain Knowledge Enhanced Graph Neural Network For PM<sub>2.5</sub> Forecasting](#)
- [Composite Neural Network: Theory and Application to PM<sub>2.5</sub> Prediction](#)
- [Reducing Effects of Swath Gaps on Unsupervised Machine Learning Models for NASA MODIS Instruments](#)
- [An improved deep learning model for predicting daily PM<sub>2.5</sub> concentration](#)
- [MetNet: A Neural Weather Model for Precipitation Forecasting](#)

## Data Sources:

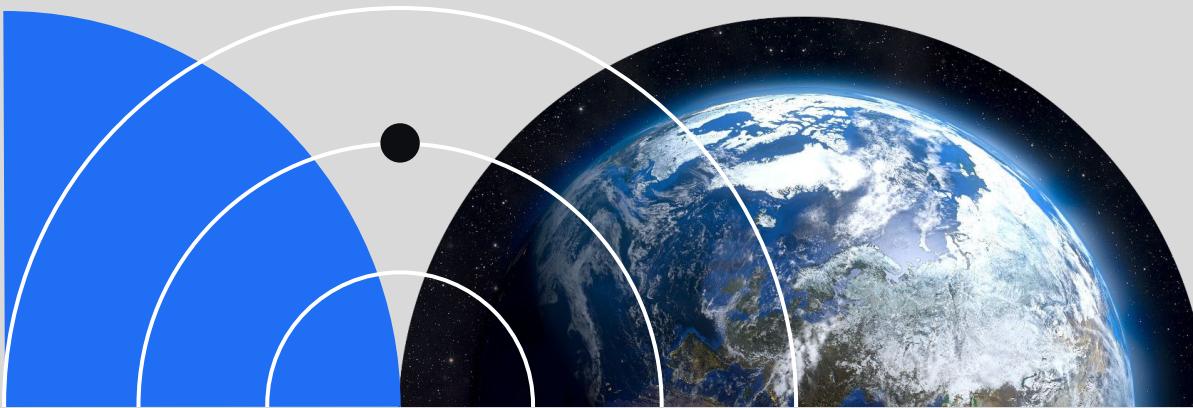
- [NASA EarthData Portal](#)
- [MAIAC](#)
- [MISR](#)
- [NASADEM](#)
- [GFS](#)
- [ECMWF](#)
- [Calipso Lidar](#)
- [GEOS Composition Forecast](#)

## Educational Resources:

- [How to Predict PM<sub>2.5</sub> Using MAIAC Aerosol Optical Depth](#)
- [GFS API Github](#)
- [NASADEM Example](#)
-



# Appendix



# Terminology

---

## PM2.5

Fine inhalable particles, with diameters that are 2.5 micrometers and smaller.

---

## Surface Level PM2.5

Our prediction target. PM2.5 at ground level, when most harmful to human health.

---

## PM2.5 Monitoring Station

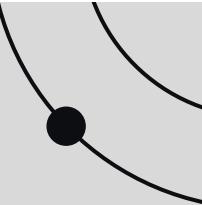
Current state for air quality reporting, but reference grade stations are costly. (\$2,000 to \$10,000+ USD<sup>1</sup>)

---

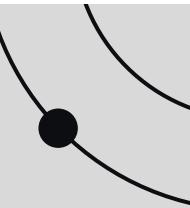
## Aerosol Optical Depth (AOD)

AOD is a satellite measure that tells us how much direct sunlight is prevented from reaching the ground by aerosol particles.

---



# Data sources



**MAIAC**

Aerosol Optical Depth

PIPELINE  
DONE

**GFS**

Weather Forecasts

PIPELINE  
DONE

**NASADEM**

Elevation  
(Topology)

IN  
PROGRESS

**Copernicus**

Proxy Aerosol Optical Depth

IN  
PROGRESS

**Goddard**

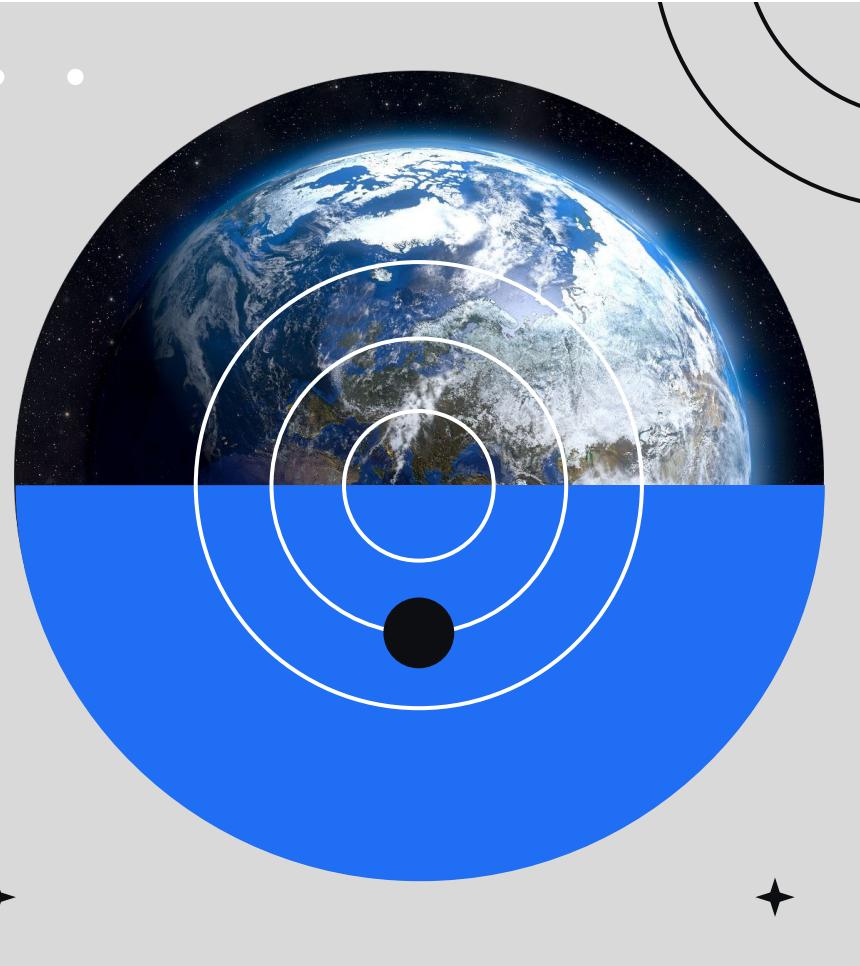
Pollutant Forecasts  
(Chemical Transport Model)

IN  
PROGRESS



*“Air pollution is one of the greatest environmental threats to human health. It can result in heart and chronic respiratory illness, cancer, and premature death.*

- ★ *Currently, no single satellite instrument provides ready-to-use, high resolution information on surface-level air pollutants. This gap in information means that millions of people cannot take daily action to protect their health.”*



# Competitive Landscape



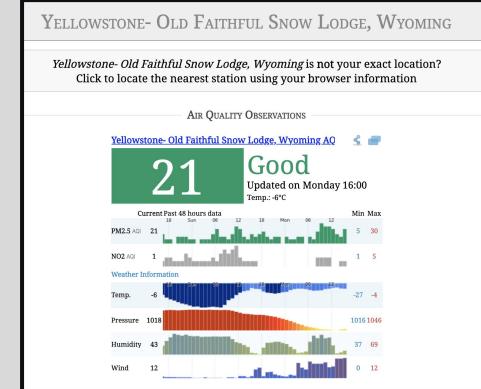
AirNow.gov

**Limitations:** Current and forecast air quality maps and data for more than 500 cities across the U.S.



IQAir.com

**Limitations:** Only ~ 75 countries with air quality reports & forecasts.

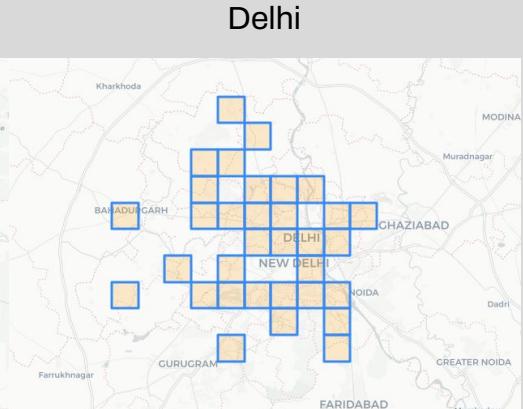
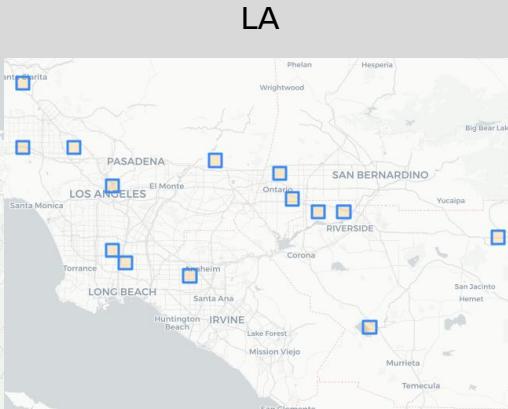
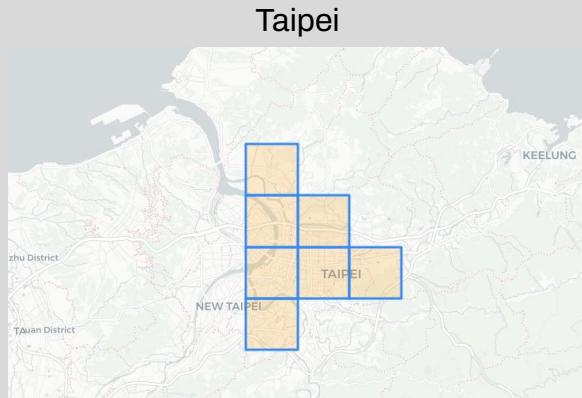


World Air Quality Forecast

**Limitations:** Relies on nearest ground station to estimate local conditions. Above example: 98 miles away.

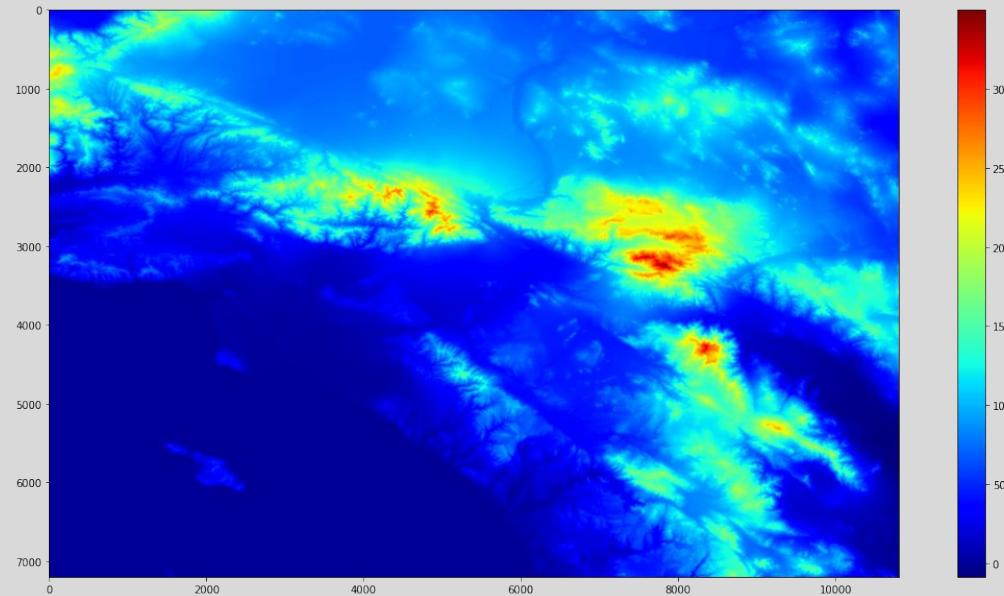
# Exploratory Data Analysis of PM2.5

Target areas for the initial model with labels



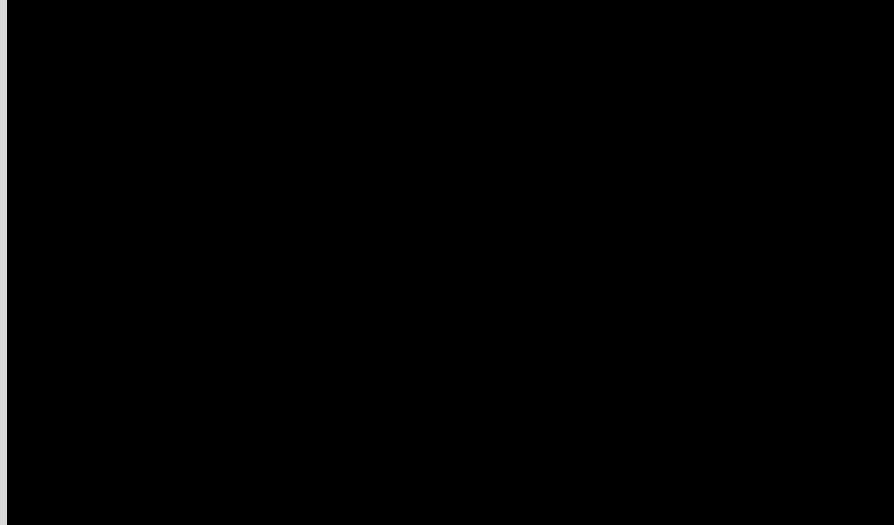
# Exploratory Data Analysis

Topography is important for identifying areas of dust particle concentrations (LA basin)

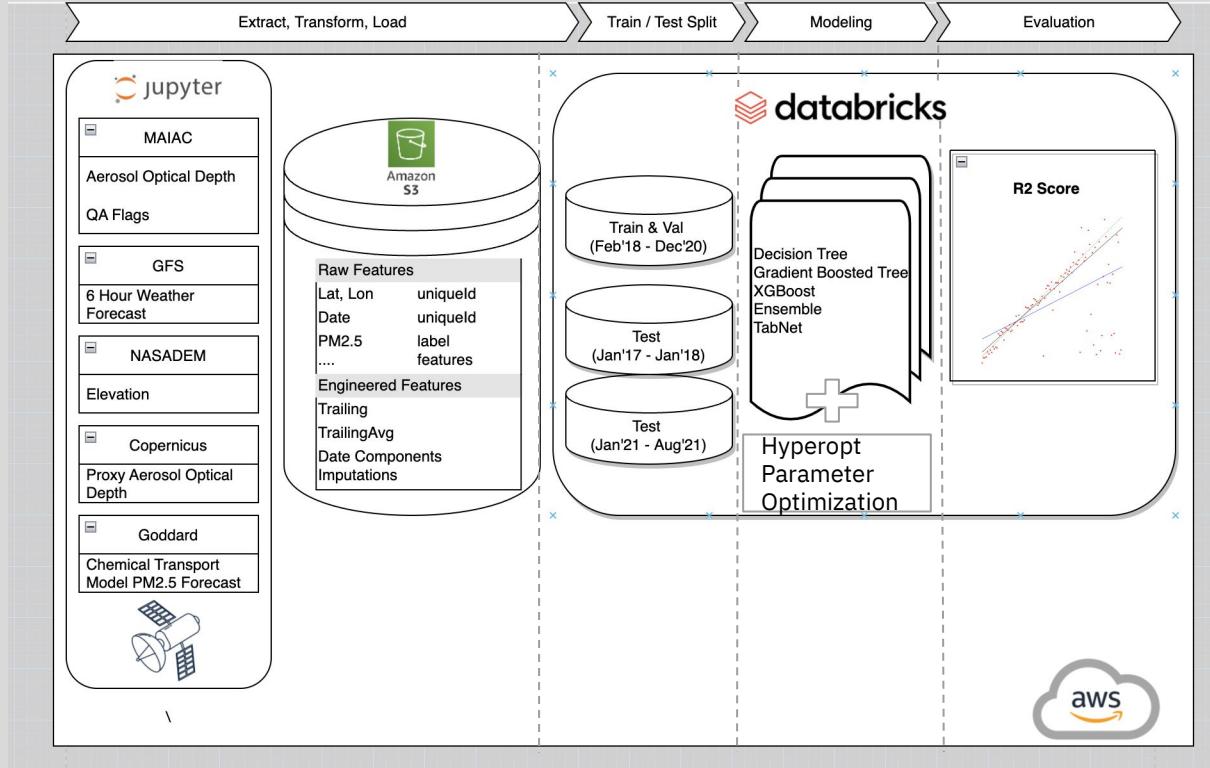


# Exploratory Data Analysis of PM2.5

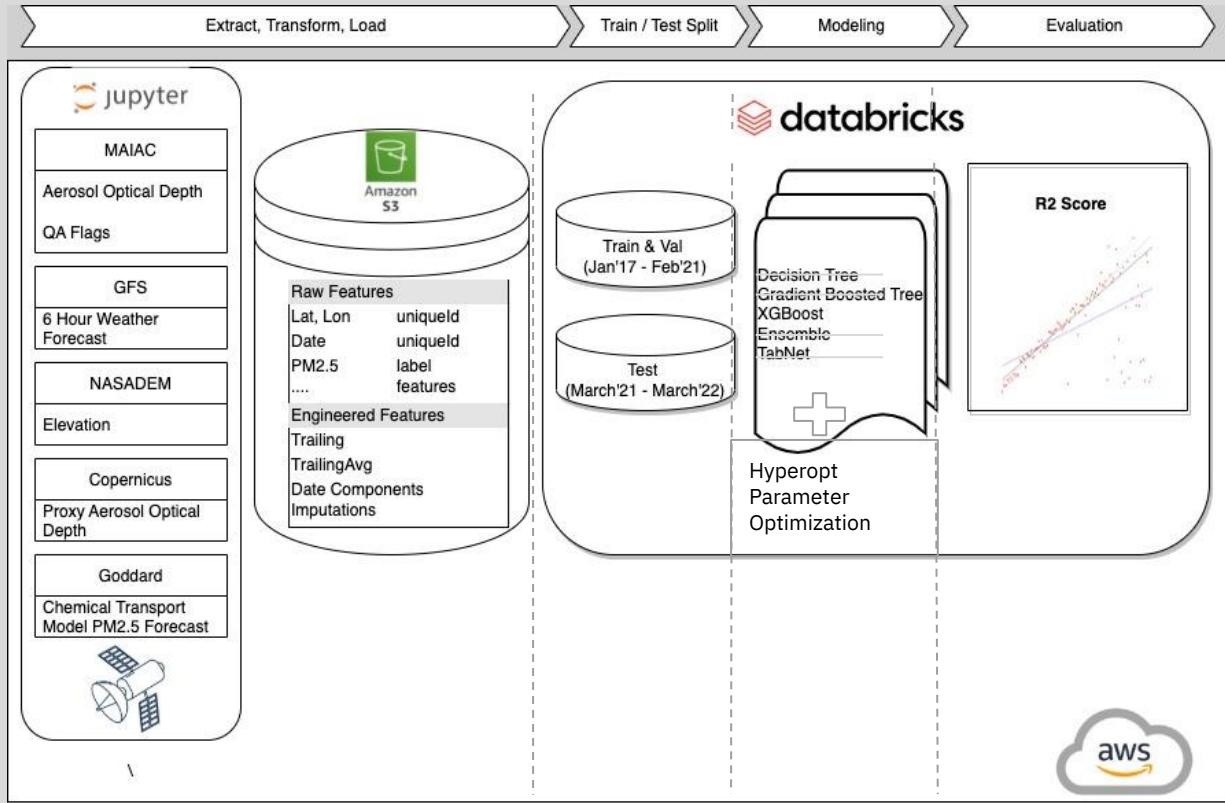
The aerosol data has missing values (example: New Delhi, January 2020)



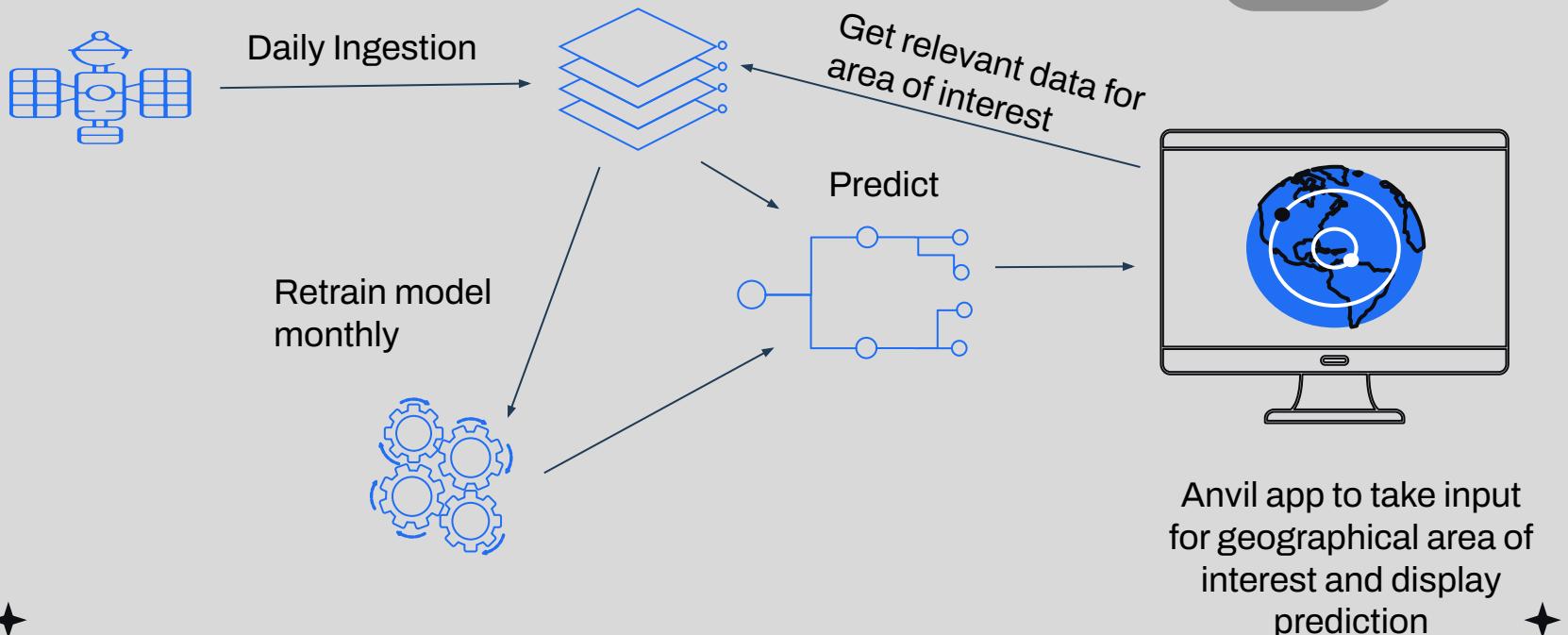
# Validation Model Architecture

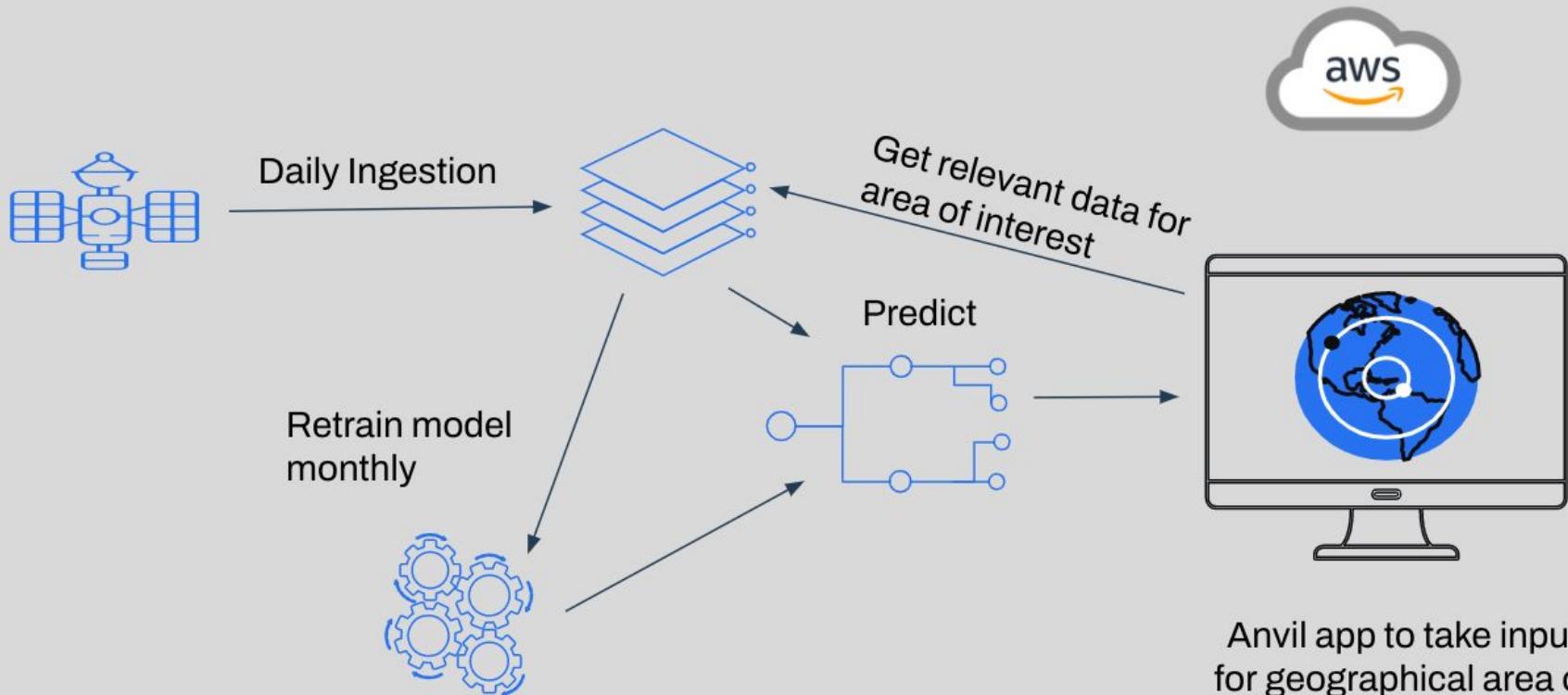


# Deployed Model Architecture



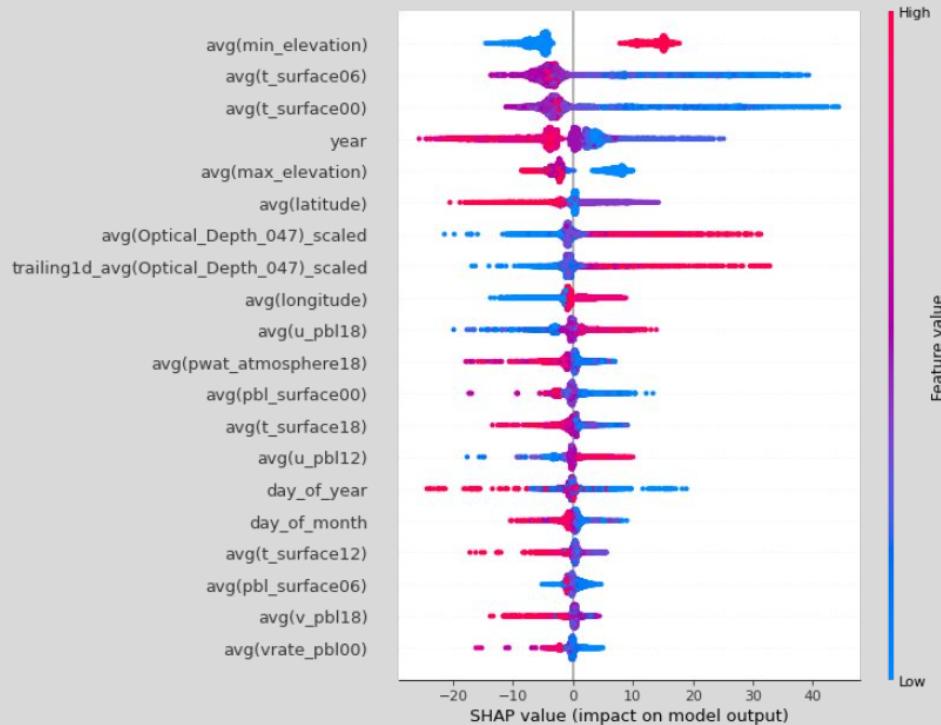
# Generating Live Predictions



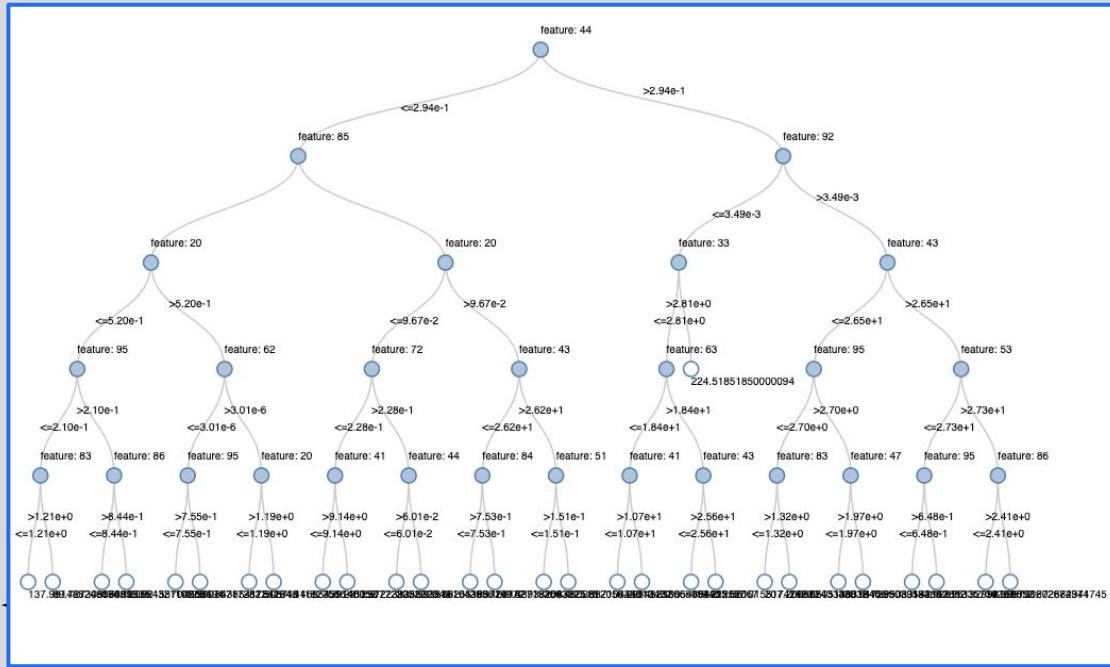


Anvil app to take input  
for geographical area of  
interest and display  
prediction

# AOD, Temp, Min Elevation, Seasonality have highest impact on PM2.5 predictions.



# PM2.5 Prediction with Tree Based Models



Feature	Importance
44 avg(pbl_surface00)	0.514245
92 change_in_t	0.128555
43 avg(t_surface00)	0.100104
85 day_of_year	0.083964
20 date_classVec	0.062820
95 rank	0.044044
53 avg(t_surface06)	0.014239
72 avg(vrate_pbl12)	0.011828
86 week_of_year	0.010578
84 day_of_week	0.007145

# AOD Imputation



## EXPLORE

Explore missing data to find patterns or randomness.



## FEATURES

Treating Imputations a separate ML problem, identifying features that would help predict the missing values.



## PREPARE

Temporary imputation for those features that had null values and feature vectorization



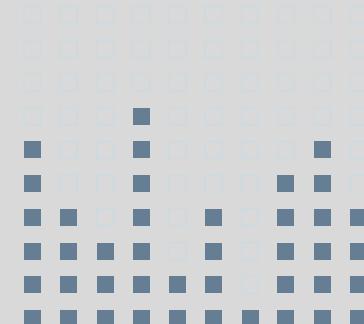
## MODEL

Impute missing values using other training data features - XGBoost worked best



## IMPUTE ALL

Run imputation sequentially so each variable benefits from previously imputed features



# Learnings



## 80% Pipelining, 20% Modeling

~80% of our time was spent setting up pipelines, extracting relevant features, cleaning data, and joining.



## Simple > Complex

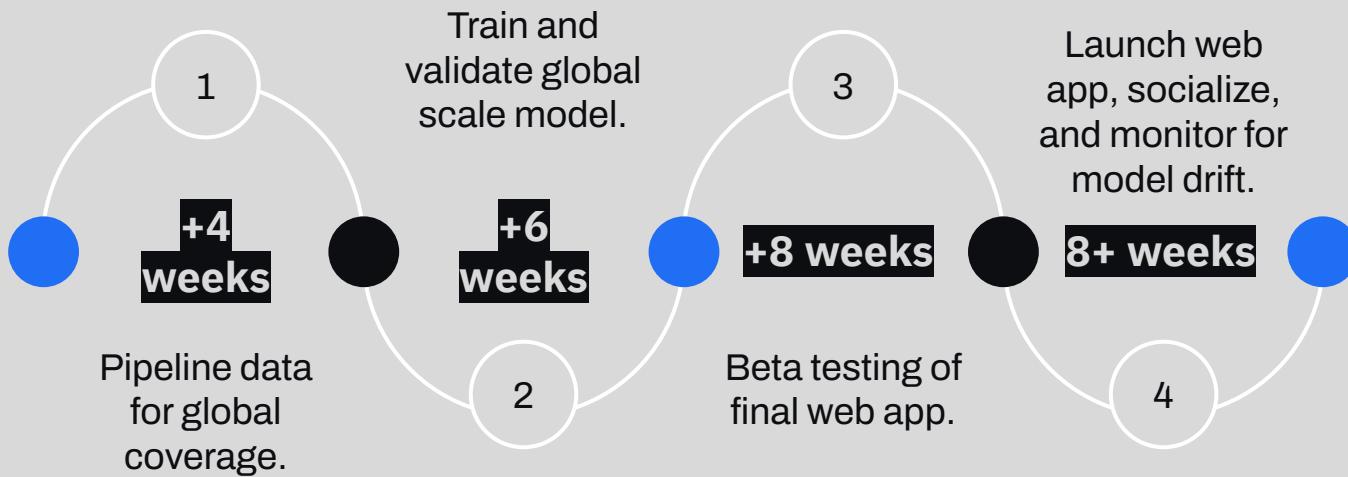
A simple, well-tuned XGBoost Model outperformed more complex approaches.



## Transferable Approach

Scalable monitoring using satellite data can be applied to predicting trace gases, snowpack levels, crop yield, environmental policy adherence, and many more critical domains.

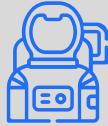
# Product Roadmap



# 1,500

# 1 of 80

# 75%



Mercury is the closest planet to the Sun and the smallest of them all



Venus has a beautiful name and is the second planet from the Sun



Despite being red, Mars is actually a cold place. It's full of iron oxide dust

# MEET OUR BEST TEACHERS

**SUSAN  
BONES**

You can speak a bit  
about this person here



**TIMMY  
JIMMY**

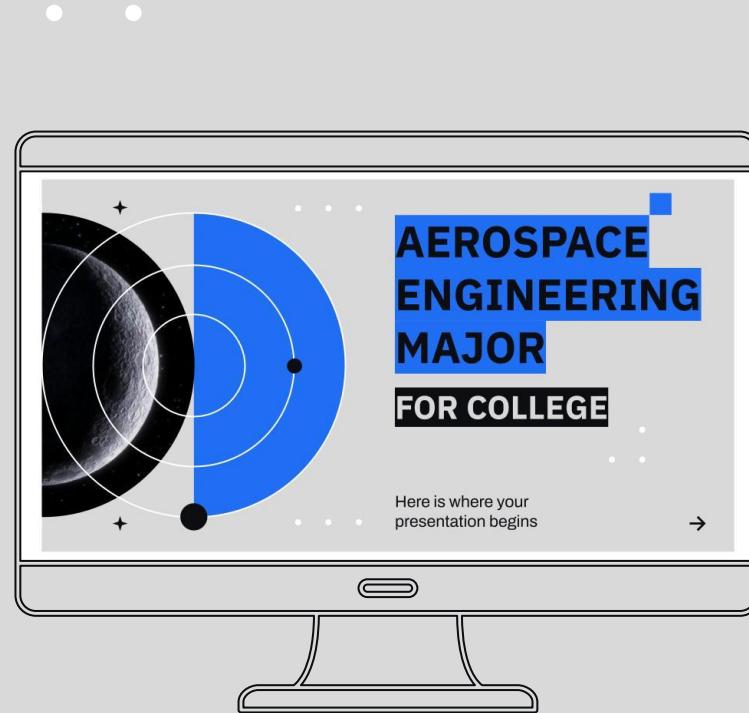
You can speak a bit  
about this person here

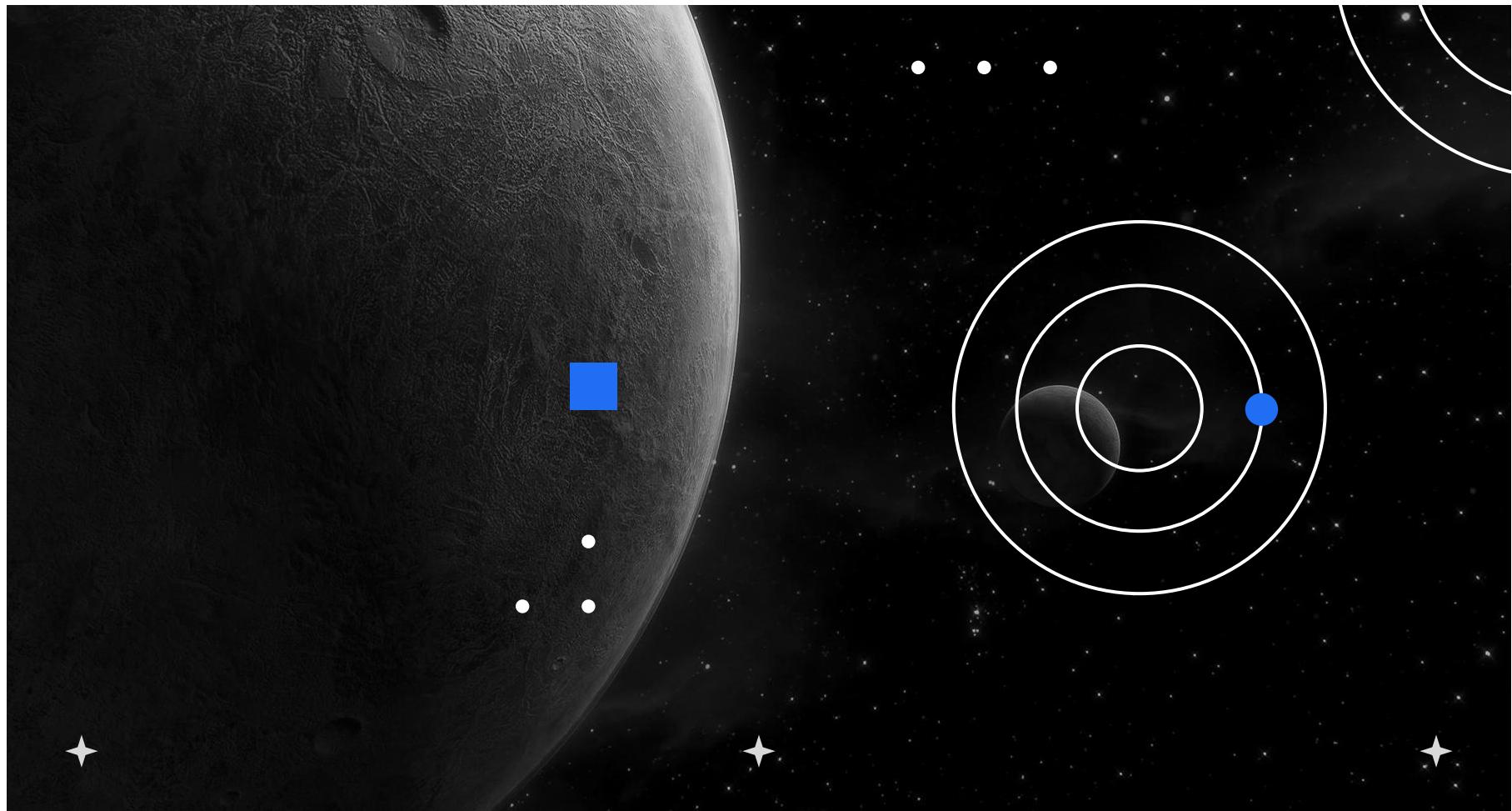


# COMPUTER MOCKUP



You can replace the image on the screen with your own work. Just right-click on it and select “Replace image”





# ORGANIZE YOUR SPACE AND TIME

**WHERE**

**WHEN**

**SUBJECT**

---

Space 1

10:00 h

Applied mathematics

---

Space 2

12:00 h

Spatial calculations

---

Space 3

16:00 h

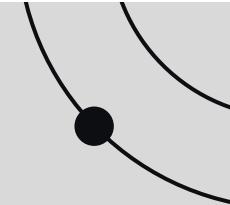
Aerospace physics

---

Space 4

18:00 h

Physical and mental preparation



# INTERNATIONAL SCHOOL



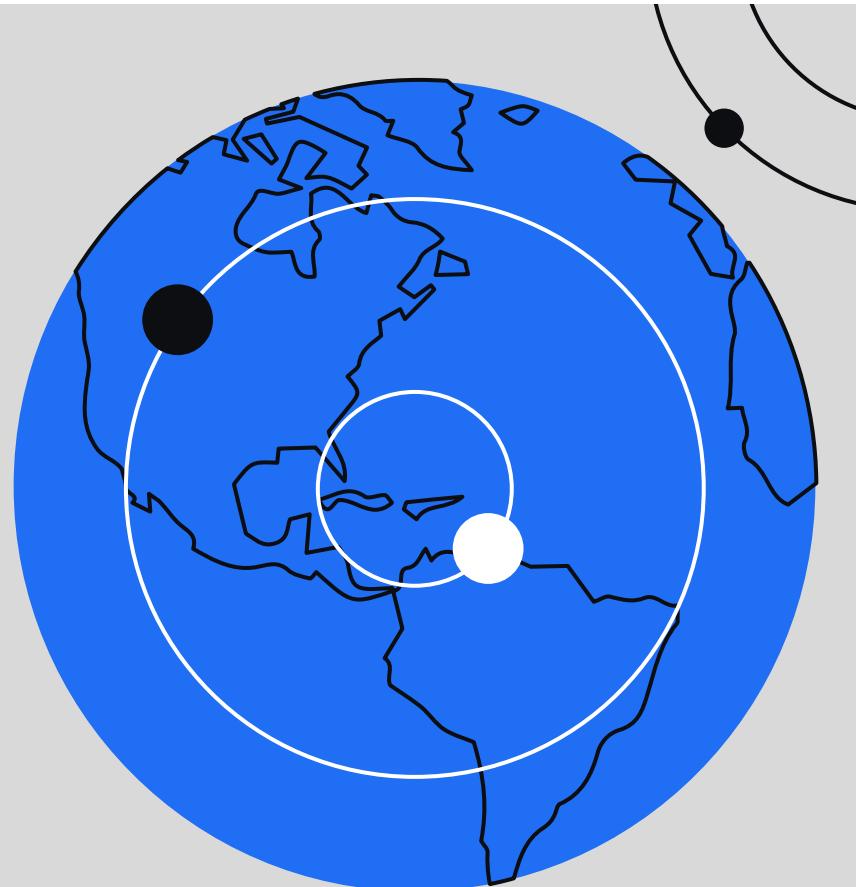
## LOCATION

Venus is the second planet from the Sun



## LOCATION

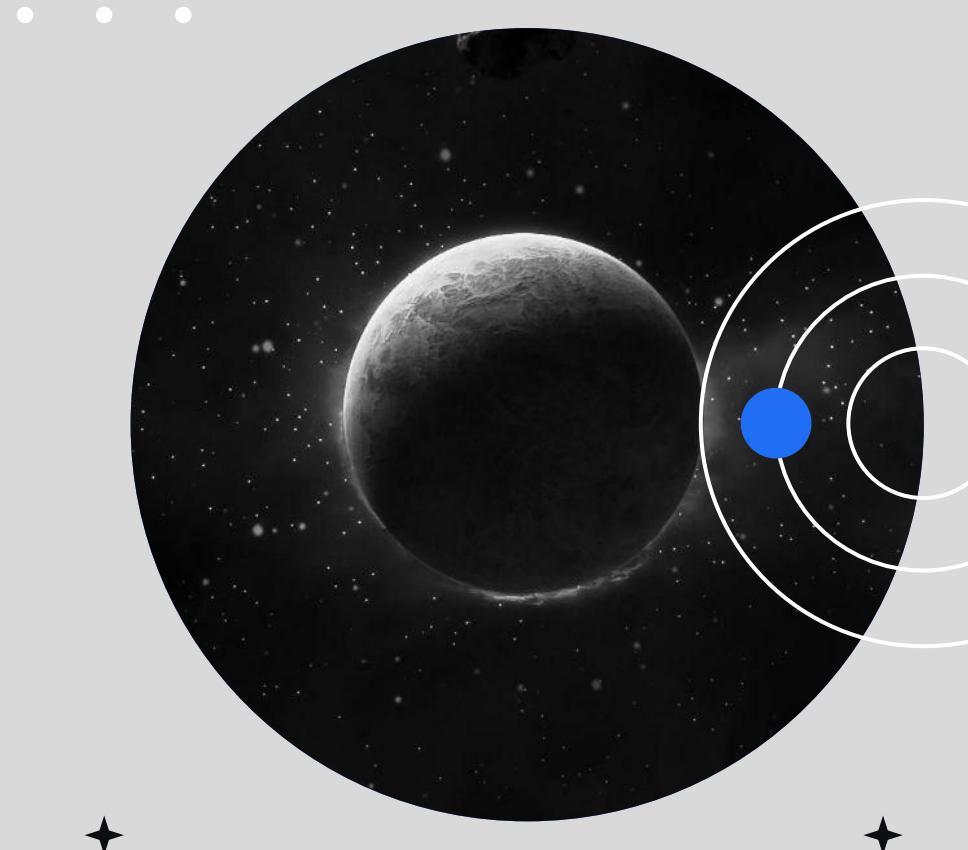
Despite being red, Mars is a cold place

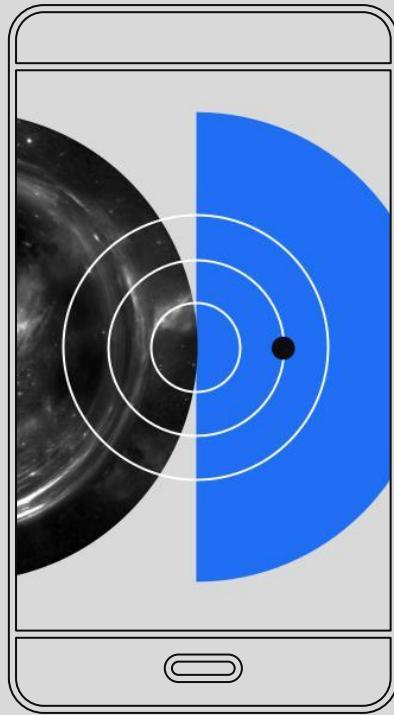


# REINFORCE THE CONCEPT



Images reveal large amounts of data, so remember: use an image instead of a long text. Your audience will appreciate it





# SMARTPHONE MOCKUP

You can replace the image on the screen with your own work. Just right-click on it and select “Replace image”

# TABLET MOCKUP



You can replace the image on the screen with your own work. Just right-click on it and select “Replace image”

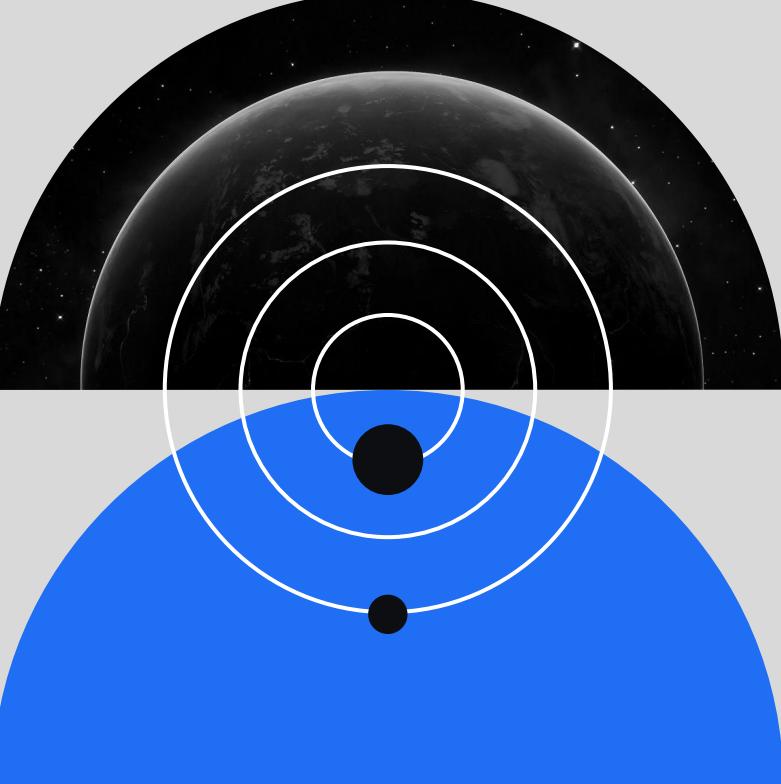


# CONTENTS OF THIS TEMPLATE

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

1. A slide structure based on a presentation for education, 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. To view this template with the correct fonts in ppt format, download and install the **fonts that we have used** on your computer. You can learn how to download and install fonts [here](#).
3. An assortment of graphic resources that are suitable for use in the presentation can be found in the **alternative resources slide**.
4. A **thanks slide**, which you must keep so that proper credits for our design are given.
5. A **resources slide**, where you'll find links to all the elements used in the template.
6. **Instructions for use**.
7. 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.



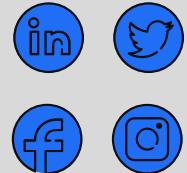
# THANKS!

**DO YOU HAVE ANY QUESTIONS?**

[youremail@freepik.com](mailto:youremail@freepik.com)

+91 620 421 838

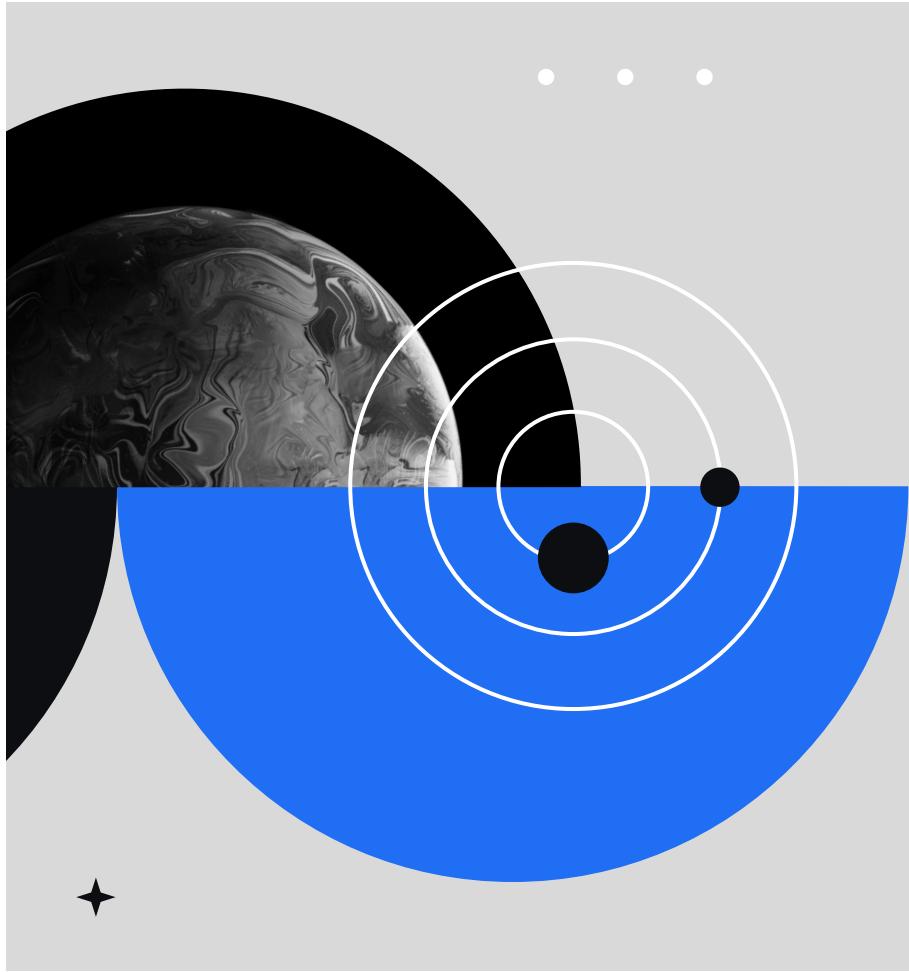
[yourcompany.com](http://yourcompany.com)



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

Please keep this slide for attribution



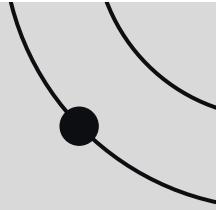


# WELCOME ABOARD

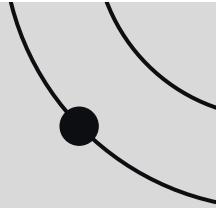
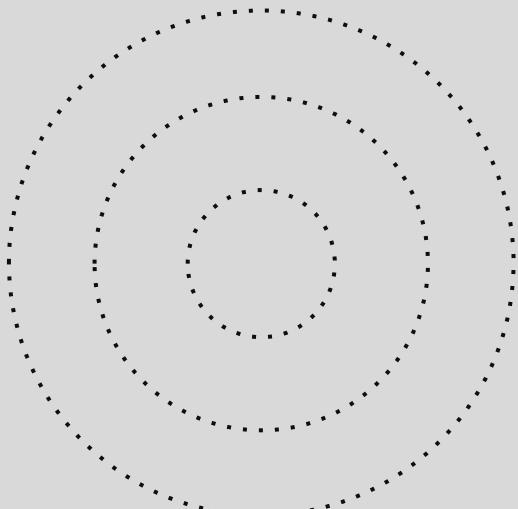
This can be the part of the presentation where you introduce yourself, write your email...



# ICON PACK: ENGINEERING



# ALTERNATIVE RESOURCES



# RESOURCES OF THIS TEMPLATE

## PHOTOS:

- Galaxy night panorama
- Abstract background with dark pink and blue sphere
- Abstract background with blue and pink sphere
- Portrait of african american model
- Galaxy night panoramic
- Galaxy night view
- Galaxy night view I
- Galaxy night panoramic I
- Confident middle aged man portrait
- Confident middle aged woman portrait

## VECTORS:

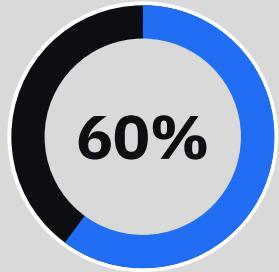
- Flat design science sale background
- Flat design science landing page template
- Flat design science facebook cover
- Flat design science facebook template

## ICON PACK:

- Icon Pack: Space Lineal

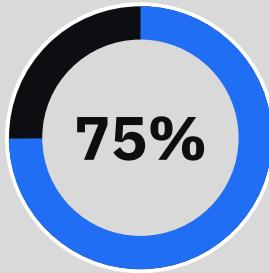


# SOME INTERESTING FACTS



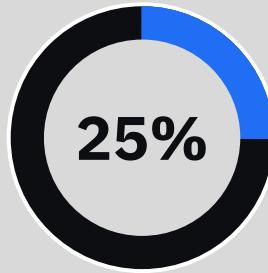
## MERCURY

It's the closest planet to the Sun



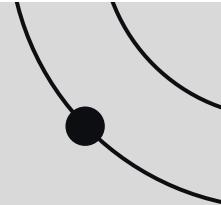
## VENUS

Venus is the second planet from the Sun

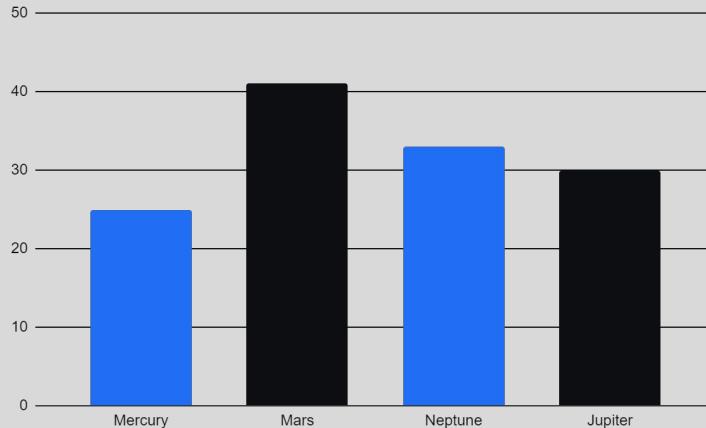


## NEPTUNE

It's the farthest planet from the Sun



# SUCCESS: STATISTICS



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

## MERCURY

It's the closest planet  
to the Sun

## MARS

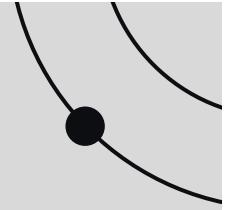
Mars is actually a very  
cold place

## NEPTUNE

It's the farthest planet  
from the Sun

## JUPITER

Jupiter is the biggest  
planet of them all



# SUCCESS RATES

**45%** MERCURY

Mercury is the closest planet to the Sun

**63%** VENUS

Venus is the second planet from the Sun

**25%** EARTH

Earth is the planet where we all live on

**40%** MARS

Despite being red, Mars is a cold place

**67%** JUPITER

It's the biggest planet in the Solar System

**10%** PLUTO

Pluto is now considered a dwarf planet



# GENDER EQUALITY

60%

MALE

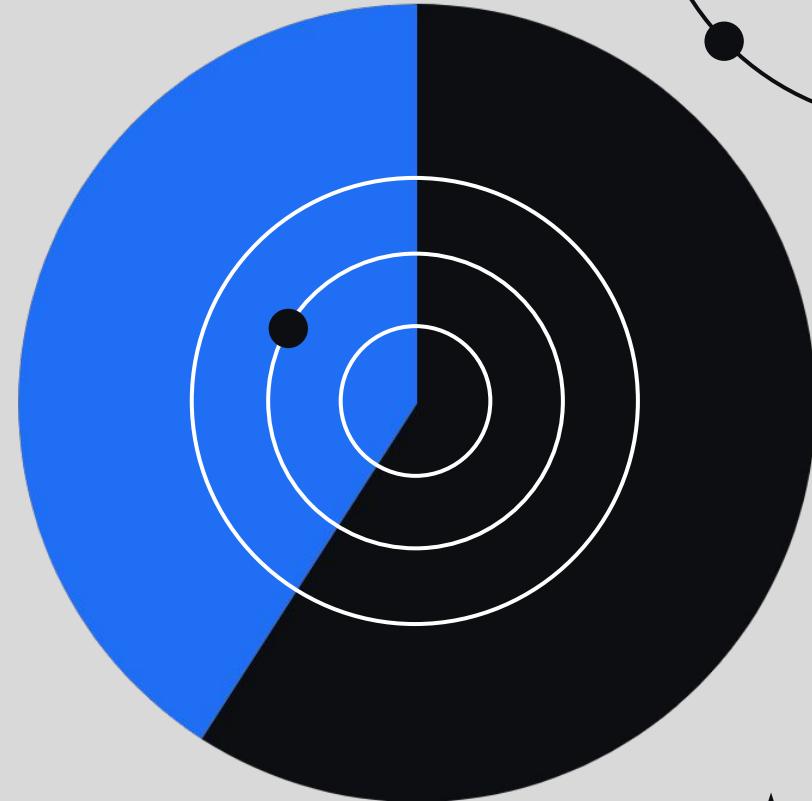
Venus is the second planet from the Sun

40%

FEMALE

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**



# FIRST SEMESTER PROGRAM



## CALCULUS I

Mercury is the closest planet to the Sun



## PHYSICS

Venus is the second planet from the Sun



## CALCULUS II

Mars is actually a very cold place



## ALGEBRA

Jupiter is the biggest planet of them all



## STATISTICS

Saturn is composed of hydrogen and helium



# Instructions for use

If you have a free account, in order to use this template, you must credit **Slidesgo** by keeping the **Thanks** slide. Please refer to the next slide to read the instructions for premium users.

## **As a Free user, 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)

As a Premium user, you can use this template without attributing Slidesgo or keeping the "Thanks" slide.

## 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:

## **IBM Plex Sans**

(<https://fonts.google.com/specimen/IBM+Plex+Sans>)

## **Archivo**

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

#0c0e11

#1f6ef3

#d9d9d9

#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



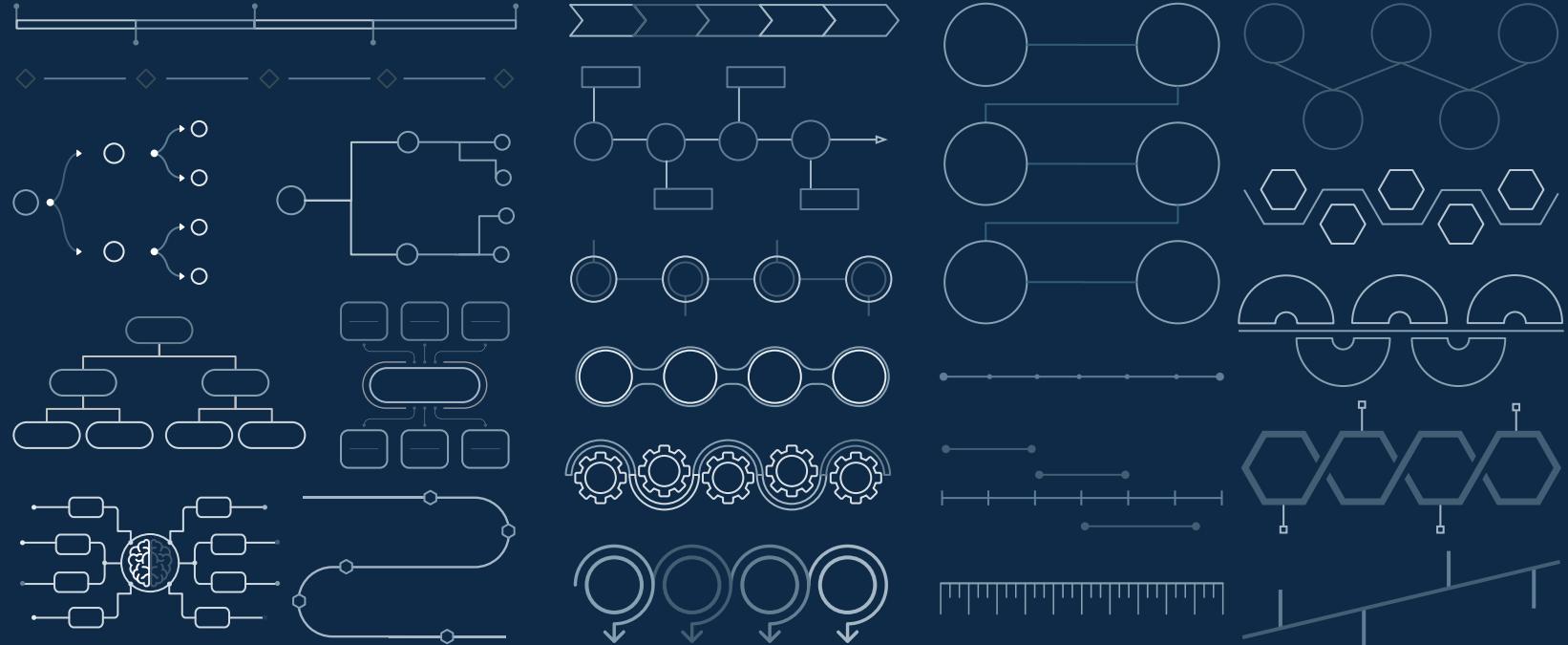
Cuate

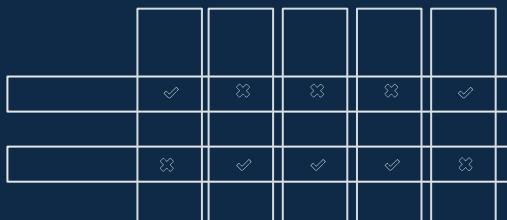
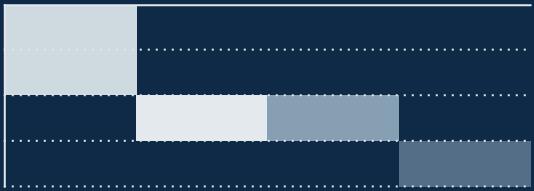
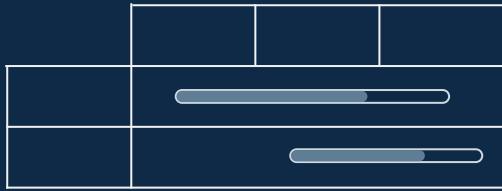
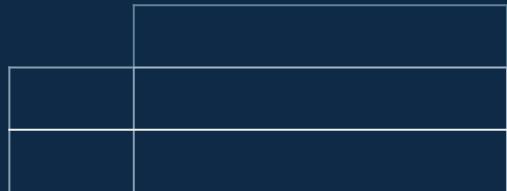
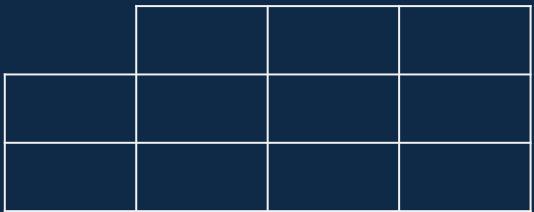
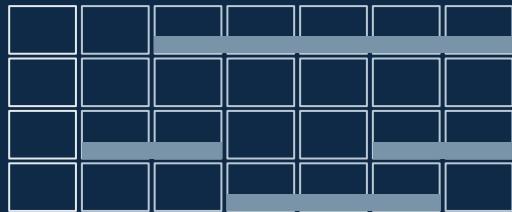
# Use our editable graphic resources...

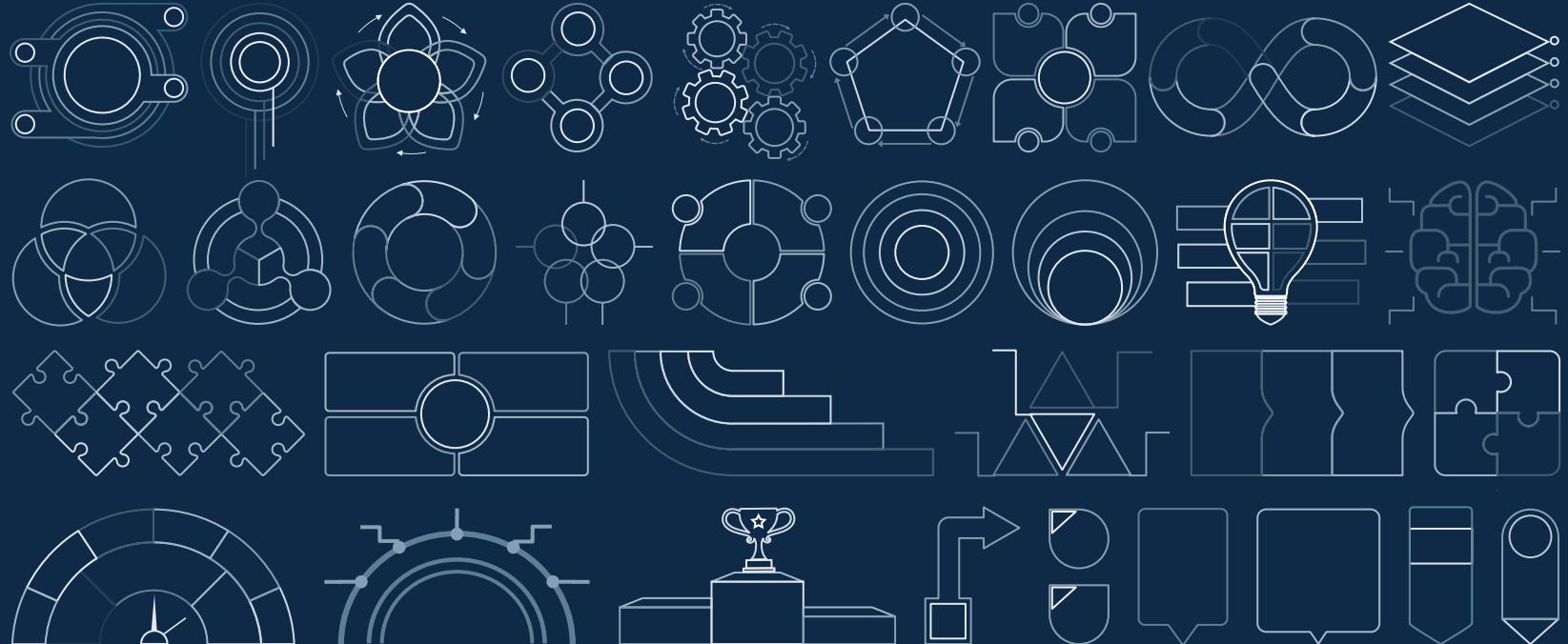
You can easily resize these resources without losing quality. To change the color, just ungroup the resource and click on the object you want to change. Then, click on the paint bucket and select the color you want. Group the resource again when you're done. You can also look for more infographics on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

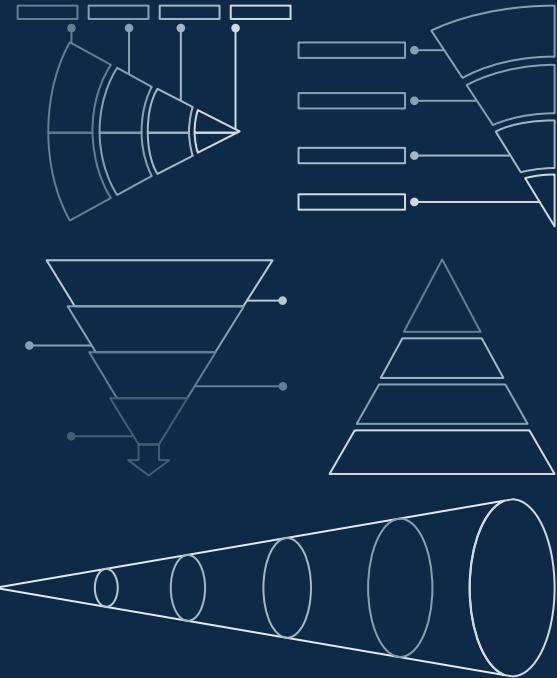
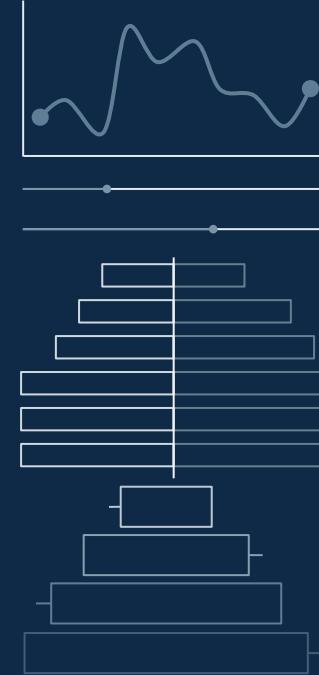
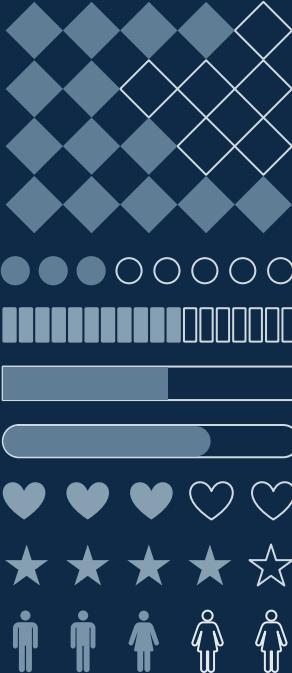
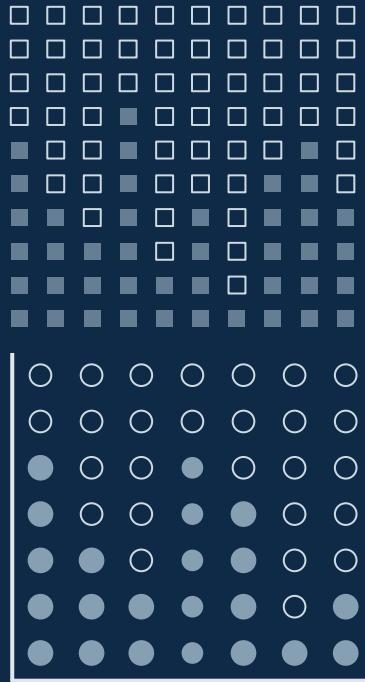












# ...and our sets of editable icons

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



