

SC1015 Mini Project

Engineer Salary Prediction

Group 4

Faiz Bin Rosli

Kang Jun Hui, Bryan

Low Zheng Han

Introduction

Engineering Career Prospect

- Many of us will be future engineers
- Form realistic expectations for our future careers

And the most important part:
It's money!



PROBLEM

How can engineering undergraduates predict their expected salary?

Does it pay well?



Is there demand for
what i'm currently
studying?



What can I expect for
various different jobs
under engineering?



PROBLEM DEFINITION

How can different variables affect the salaries of engineering students' after graduation?



VARIABLES INVOLVED

GENDER



GRADE 10/12 EXAM MARKS

COLLEGE NAMES & TIERS



DEGREE & SPECIALISATION

AMCAT* SCORES



PERSONALITY

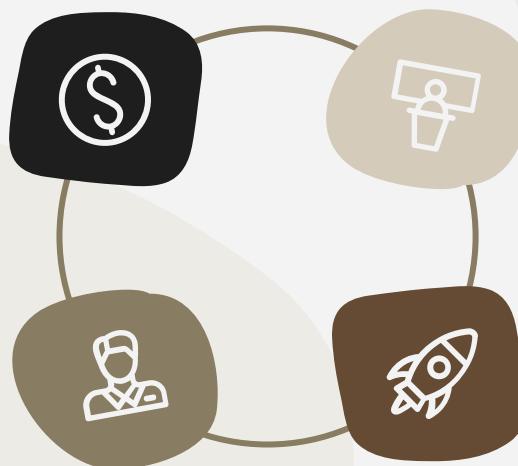
EXPLORATORY DATA ANALYSIS (EDA)

GRADE 10/12
EXAM MARKS

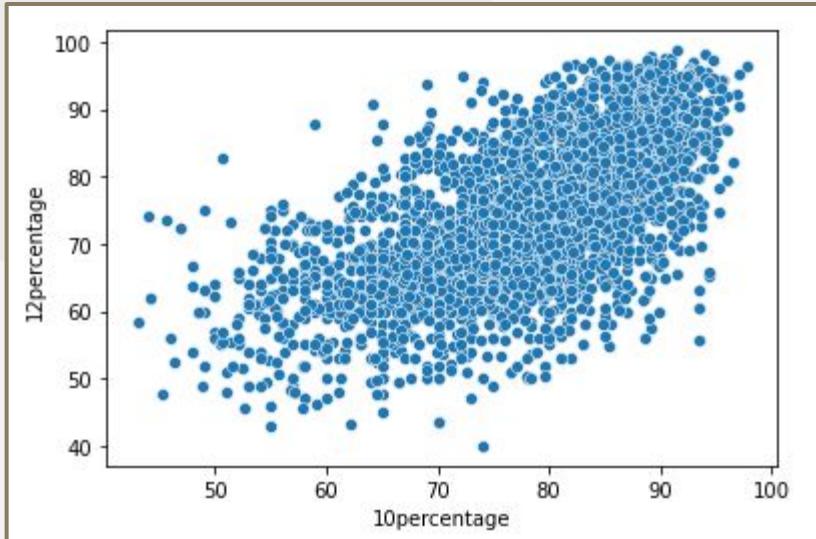
PERSONALITY

AMCAT SCORES

DEGREE &
SPECIALISATION



EDA1: GRADE 10/12 EXAM MARKS



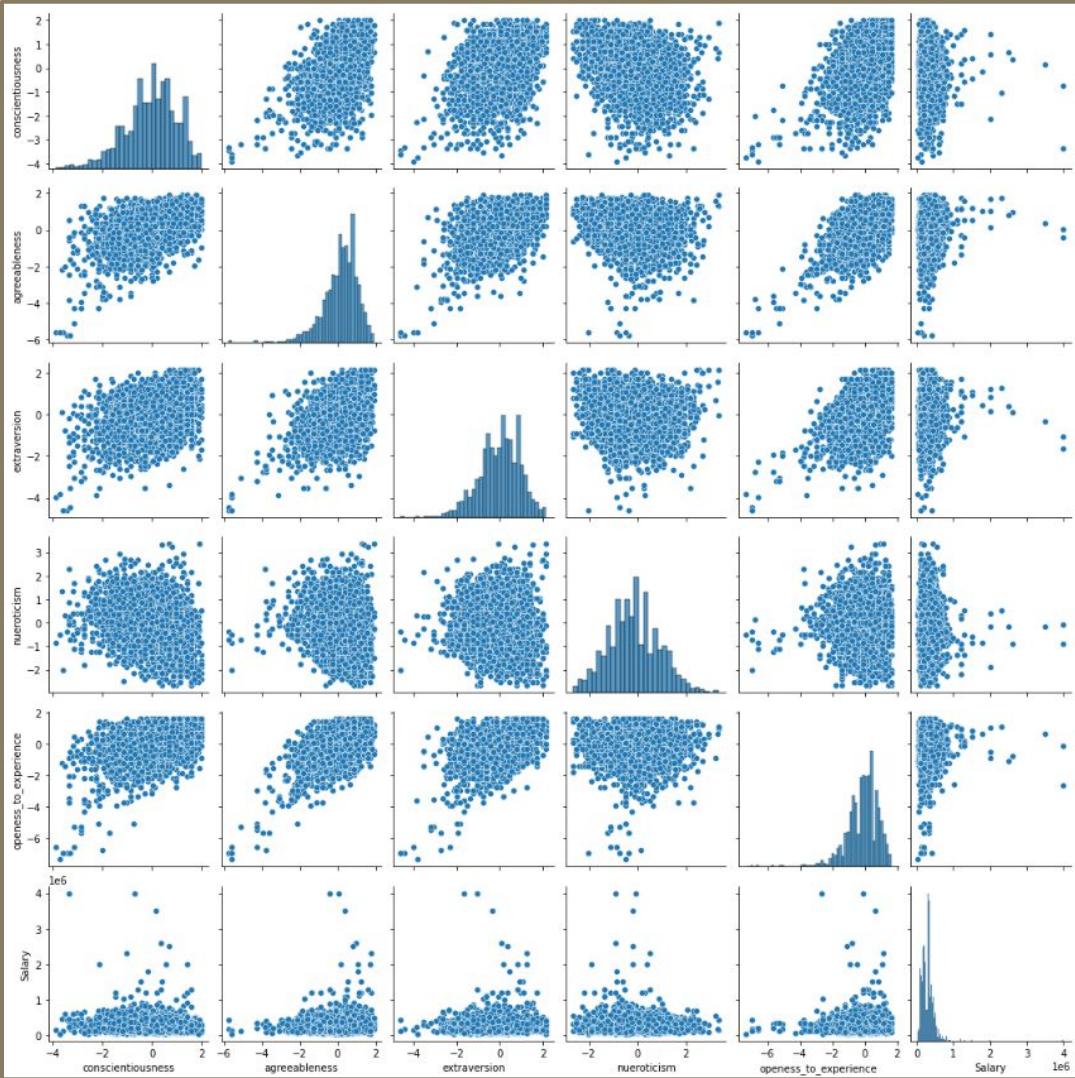
BIG CLUSTER



LINEAR RELATIONSHIP



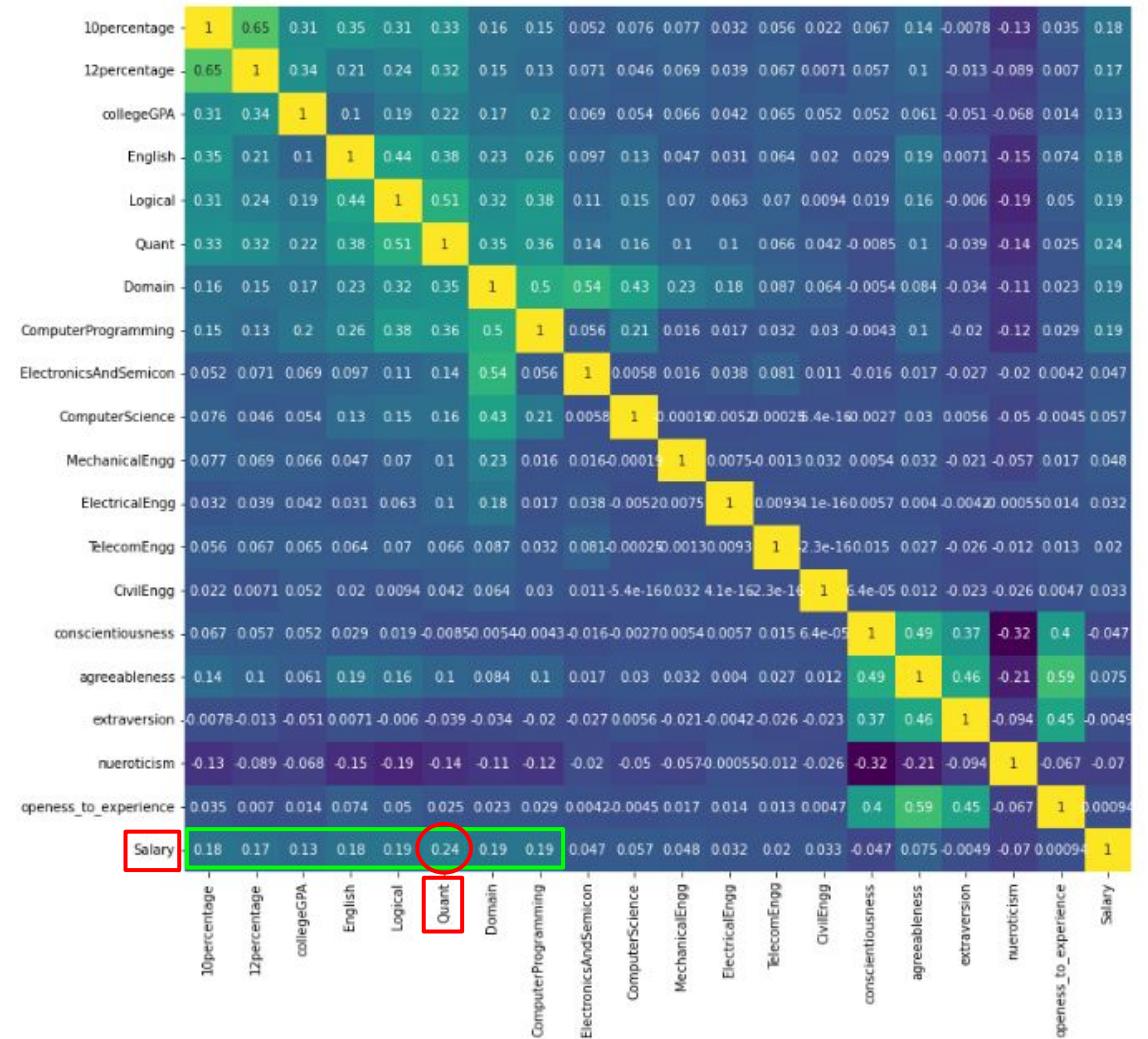
If one performs well in Grade 10, he/she should likewise perform well in Grade 12



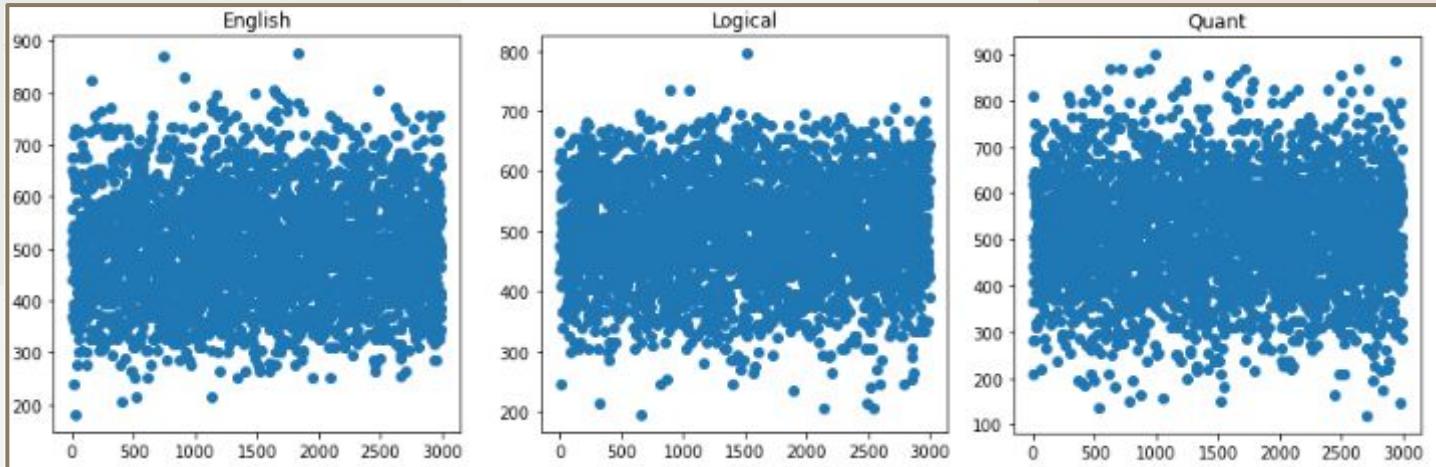
EDA2: PERSONALITY

PAIRPLOT

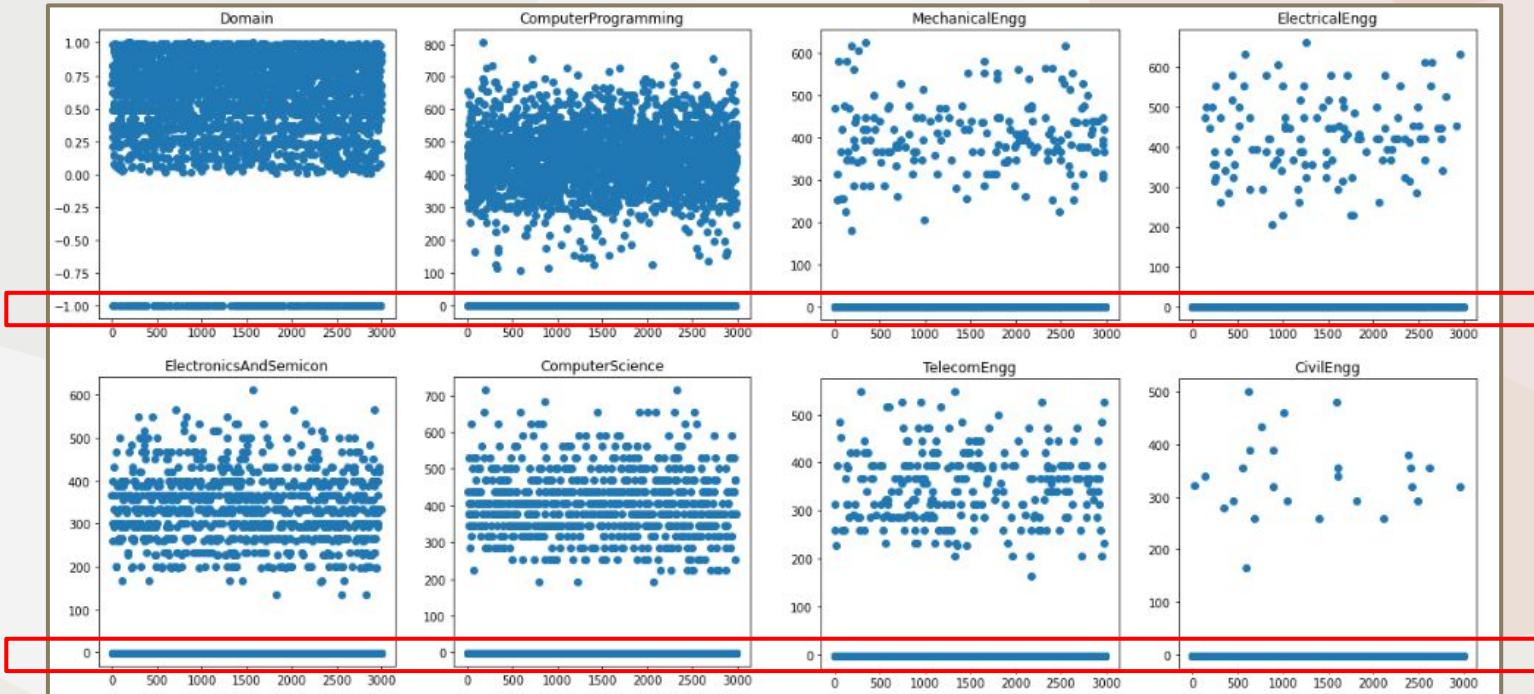
No clear relationship
between Personality
scores and Salary



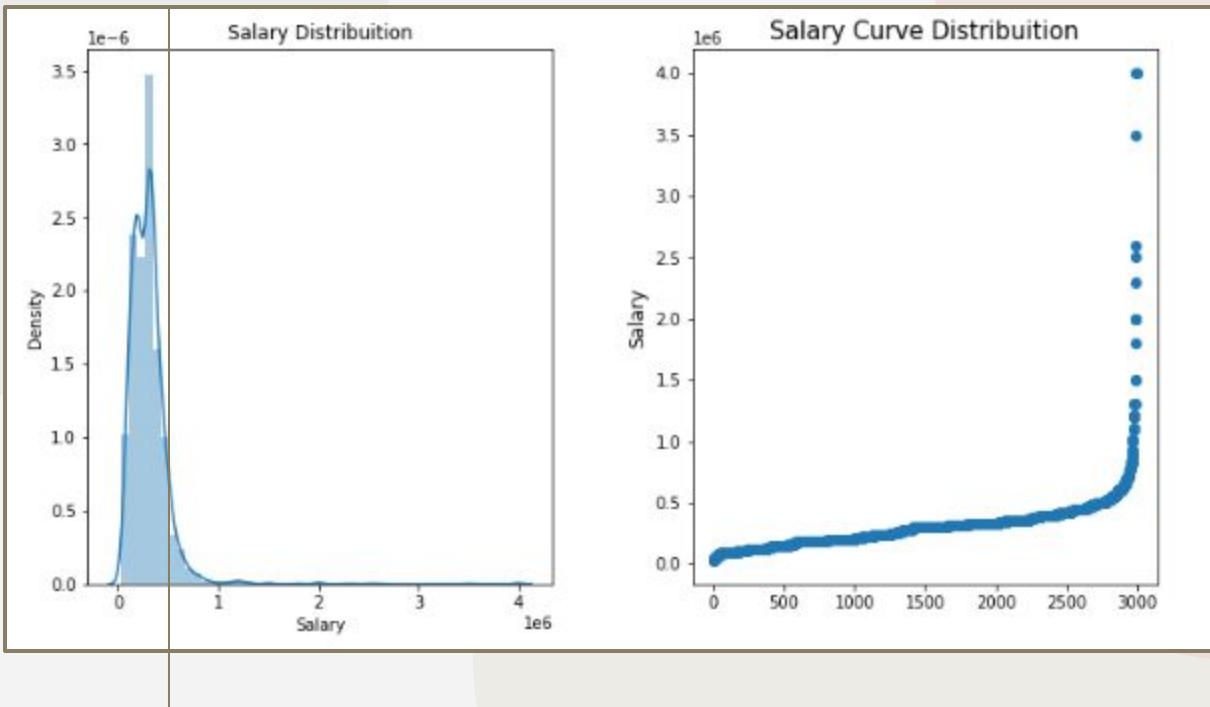
EDA3: AMCAT SCORES



EDA3: AMCAT SCORES

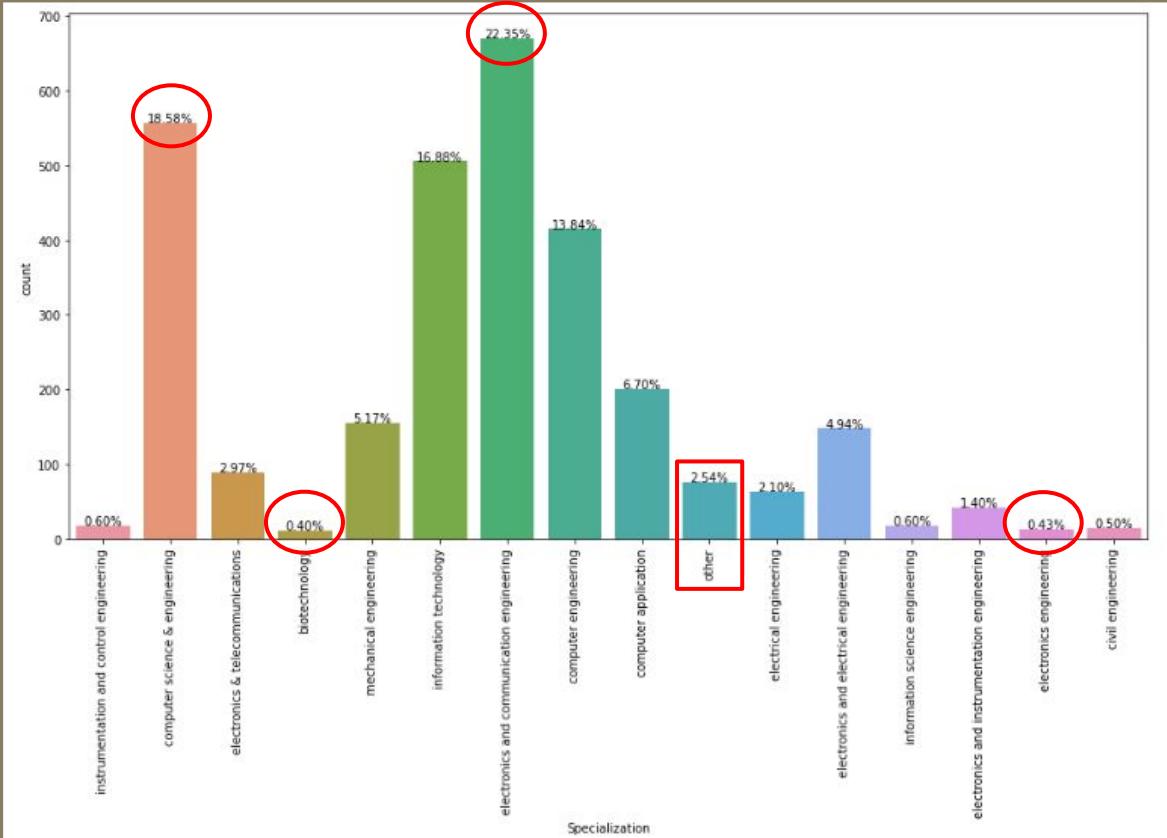


SALARY DISTRIBUTION



INR 500,000 \approx SGD 8950

EDA4: DEGREE & SPECIALISATION



HIGHEST COUNT

Electronics & Communication Eng.



2ND HIGHEST COUNT

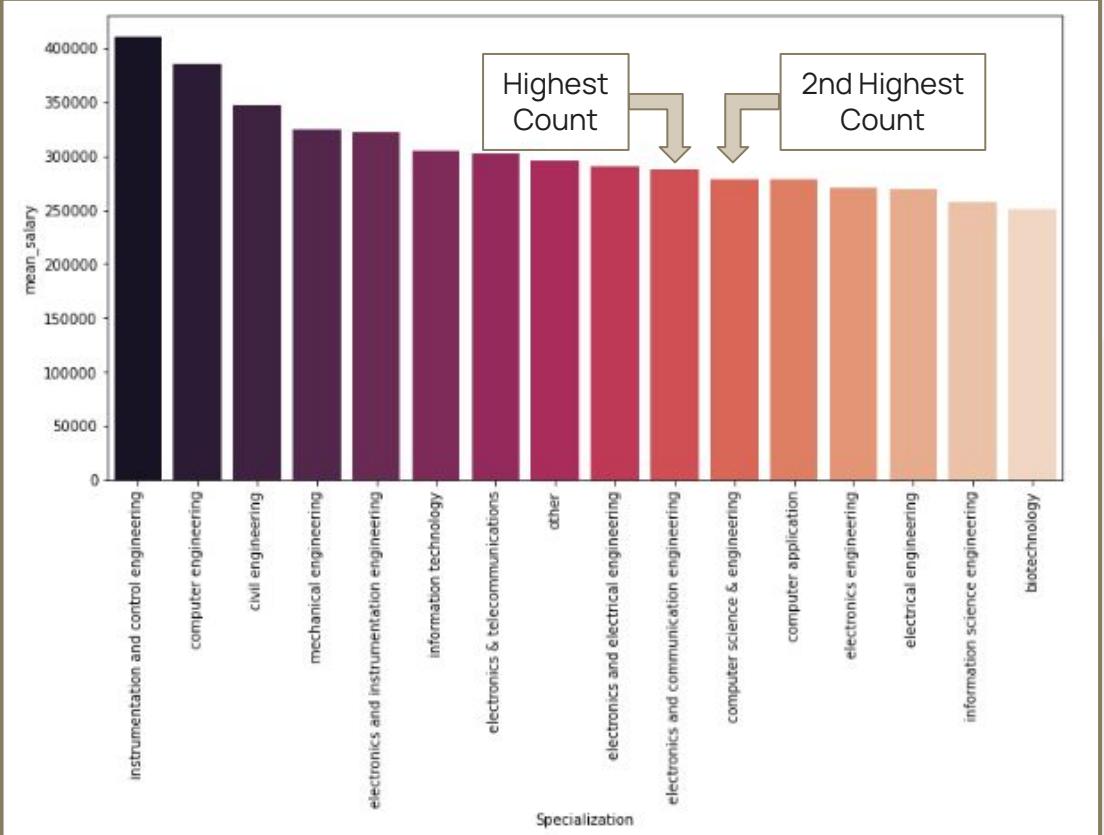
Computer Science & Eng.



UNPOPULAR

Biotechnology
Electronics Eng.

EDA4: DEGREE & SPECIALISATION



HIGHEST SALARY

Instrumentation & Control Eng.



LOWEST SALARY

Biotechnology

DATA PREPARATION



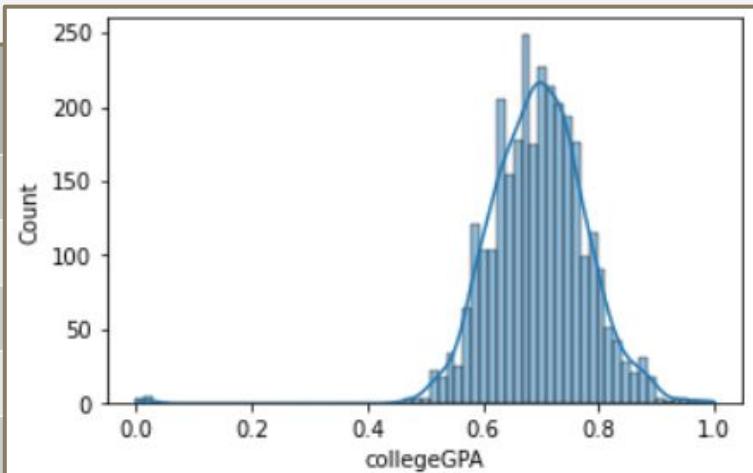
ONE-HOT ENCODING

Representation of Categorical Variables as Binary Vectors

Index	Animal	One-Hot code	Index	Dog	Cat
0	Dog		0	1	0
1	Cat		1	0	1
2	Sheep		2	0	0
3	Horse		3	0	0
4	Lion		4	0	0

MIN-MAX SCALER

Feature-Wise Normalisation



MACHINE LEARNING

- Ordinary Least Squares
- XGBoost
- AdaBoost
- Gradient Boosting



Root Mean Square Error (RMSE)
&
Mean Absolute Percentage Error (MAPE)

ORDINARY LEAST SQUARES REGRESSION

Split the dataset into Train and Test sets uniformly at random

Build a Linear Model

Check Multicollinearity between the Variables

- Calculate Variance Inflation Factor (VIF)
- $VIF = 1/(1-R^2)$

Further Data Cleaning

- Drop variables with high p-value & VIF
- Repeat until every column's p-value < 0.005 & VIF < 5

Features	VIF
computer application	inf
MCA	inf
B.Tech/B.E.	780.42
CivilEngg	456.58
MechanicalEngg	130.24
ElectricalEngg	122.47
collegeGPA	84.01
TelecomEngg	63.43
electronics and communication engineering	46.67
ElectronicsAndSemicon	41.62
computer science & engineering	38.06
ComputerScience	37.51
information technology	33.86
ComputerProgramming	30.73
computer engineering	28.46
10percentage	26.43
Logical	21.50
12percentage	20.10
Domain	19.97
Quant	18.45
English	14.76
mechanical engineering	12.36
electronics and electrical engineering	12.07
M.Tech/M.E	10.50
electronics & telecommunications	6.70
other	5.41
electrical engineering	4.64
electronics and instrumentation engineering	3.97
civil engineering	2.35
instrumentation and control engineering	2.26
agreeableness	2.14
information science engineering	1.98
electronics engineering	1.97
openness_to_experience	1.81
conscientiousness	1.57
extraversion	1.45
neuroticism	1.21

OLS Regression Results						
Dep. Variable:	Salary	R-squared:	0.121			
Model:	OLS	Adj. R-squared:	0.106			
Method:	Least Squares	F-statistic:	8.102			
Date:	Mon, 18 Apr 2022	Prob (F-statistic):	1.16e-37			
Time:	16:22:32	Log-Likelihood:	3252.9 <th></th> <th></th> <th></th>			
No. Observations:	2098	AIC:	-6434.			
Df Residuals:	2062	BIC:	-6231.			
Df Model:	35					
Covariance Type:	nonrobust					
	coef	std err	t	P> t	[0.025	0.975
const	0.0318	0.068	0.470	0.638	-0.181	0.161
12percentage	0.0200	0.007	2.862	0.004	0.006	0.034
collegeGPA	0.0221	0.015	1.499	0.134	-0.007	0.053
English	0.0246	0.009	2.810	0.005	0.007	0.042
Logical	0.0078	0.010	0.782	0.434	-0.012	0.027
Quant	0.0449	0.009	4.881	0.000	0.027	0.065
Domain	0.0229	0.008	2.845	0.004	0.007	0.039
ComputerProgramming	0.0041	0.013	0.327	0.744	-0.020	0.029
ElectronicsAndSemicon	-0.0234	0.017	-1.352	0.177	-0.057	0.019
ComputerScience	-0.0188	0.017	-1.108	0.268	-0.052	0.014
MechanicalEngg	-0.0089	0.027	-0.333	0.739	-0.061	0.043
ElectricalEngg	-0.0128	0.027	-0.477	0.633	-0.065	0.040
TelecomEngg	-0.0077	0.019	-0.412	0.680	-0.044	0.029
CivilEngg	0.0434	0.051	0.849	0.396	-0.057	0.143
conscientiousness	-0.0051	0.001	-3.717	0.000	-0.008	-0.001
agreeableness	0.0046	0.002	2.676	0.008	0.001	0.008
extraversion	0.0025	0.001	1.785	0.074	-0.000	0.008
neuroticism	-0.0030	0.001	-2.404	0.016	-0.005	-0.001
openness_to_experience	-0.0022	0.001	-1.525	0.127	-0.005	0.000
B.Tech/B.E.	-0.0534	0.053	-1.013	0.311	-0.157	0.056
M.Tech./M.E.	-0.0549	0.054	-1.024	0.306	-0.160	0.056
MCA	-0.0172	0.028	-0.621	0.534	-0.071	0.037
civil engineering	0.0331	0.022	1.493	0.136	-0.010	0.077
computer application	-0.0172	0.028	-0.621	0.534	-0.071	0.037
computer engineering	0.0337	0.017	2.015	0.044	0.001	0.067
computer science & engineering	0.0097	0.017	0.585	0.559	-0.023	0.042
electrical engineering	0.0102	0.018	0.553	0.581	-0.026	0.044
electronics & telecommunications	0.0157	0.018	0.881	0.379	-0.019	0.053
electronics and communication engineering	0.0106	0.017	0.635	0.526	-0.022	0.043
electronics and electrical engineering	0.0141	0.017	0.822	0.411	-0.020	0.049
electronics and instrumentation engineering	0.0254	0.019	1.341	0.180	-0.012	0.061
electronics engineering	0.0140	0.023	0.599	0.549	-0.032	0.060
information science engineering	0.0155	0.023	0.668	0.504	-0.030	0.061
information technology	0.0151	0.017	0.903	0.367	-0.018	0.049
instrumentation and control engineering	0.0343	0.022	1.567	0.117	-0.009	0.077
mechanical engineering	0.0249	0.017	1.455	0.146	-0.009	0.059
other	0.0137	0.018	0.756	0.450	-0.022	0.049
=====						
Omnibus:	2837.681	Durbin-Watson:	2.039			
Prob(Omnibus):	0.000	Jarque-Bera (JB):	873116.225			
Skew:	7.424	Prob(JB):	0.00			
Kurtosis:	101.831	Cond. No.	1.11e+16			
=====						
Notes:						
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.						
[2] The smallest eigenvalue is 9.54e-29. This might indicate that there are strong multicollinearity problems or that the design matrix is singular.						

ORDINARY LEAST SQUARES REGRESSION

Check if Error Terms are Normally Distributed

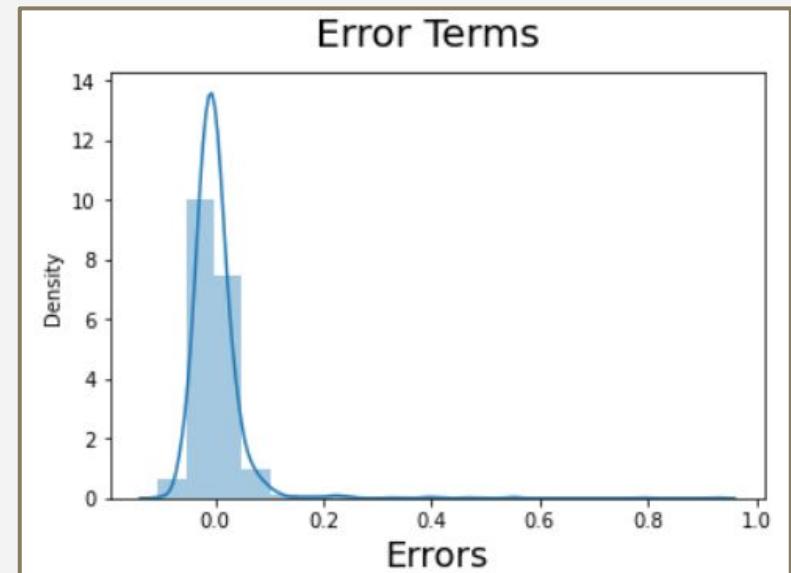
- Residual Analysis
- At a particular x-value, Error-Term = $y_{\text{actual}} - y_{\text{predicted}}$

Plot Error Terms' Histogram

R^2 Scores

- Test = 0.1186017893464345 ≈ 0.12
- Train = 0.12090024850401904 ≈ 0.12

Explained Variance ≈ 0.12



ORDINARY LEAST SQUARES REGRESSION

```
testPredDf = pd.DataFrame()  
testPredDf["Prediction"] = lr_1.predict(X_test)  
testPredDf["Actual"] = y_test.values  
testPredDf.head()
```

	Prediction	Actual
1271	0.042439	0.293821
699	0.063555	0.041614
356	0.088124	0.132409
2776	0.065689	0.075662
561	0.042133	0.041614



Root Mean Square Error (RMSE)
= 0.04751822561000885



Mean Absolute Percentage Error (MAPE)
= 73.22609412169932%

MACHINE LEARNING

- Ordinary Least Squares
- XGBoost
- AdaBoost
- Gradient Boosting



Root Mean Square Error (RMSE)
&
Mean Absolute Percentage Error (MAPE)

XGBOOST

```
testPredDf = pd.DataFrame()  
testPredDf["Prediction"] = XG_raw.predict(X_test)  
testPredDf["Actual"] = y_test.values  
testPredDf.head()
```

	Prediction	Actual
0	0.061913	0.293821
1	0.040440	0.041614
2	0.076312	0.132409
3	0.072378	0.075662
4	0.037281	0.041614



Root Mean Square Error (RMSE)
= 0.05095205901355028



Mean Absolute Percentage Error (MAPE)
= 78.9181133905711%

MACHINE LEARNING

- Ordinary Least Squares
- XGBoost
- AdaBoost ◀
- Gradient Boosting

Root Mean Square Error (RMSE)
&
Mean Absolute Percentage Error (MAPE)

ADABOOST

```
testPredDf = pd.DataFrame()  
testPredDf["Prediction"] = Ada_raw.predict(X_test)  
testPredDf["Actual"] = y_test.values  
testPredDf.head()
```

	Prediction	Actual
0	0.075709	0.293821
1	0.075709	0.041614
2	0.075709	0.132409
3	0.090674	0.075662
4	0.075709	0.041614



Root Mean Square Error (RMSE)
= 0.05161626997678874



Mean Absolute Percentage Error (MAPE)
= 105.87927813164337%

MACHINE LEARNING

- Ordinary Least Squares
- XGBoost
- AdaBoost
- Gradient Boosting



Root Mean Square Error (RMSE)
&
Mean Absolute Percentage Error (MAPE)

GRADIENT BOOSTING

```
testPredDf = pd.DataFrame()  
testPredDf["Prediction"] = GD_raw.predict(X_test)  
testPredDf["Actual"] = y_test.values  
testPredDf.head()
```

	Prediction	Actual
0	0.050754	0.293821
1	0.057460	0.041614
2	0.081150	0.132409
3	0.074732	0.075662
4	0.046603	0.041614



Root Mean Square Error (RMSE)
= 0.04789611561372218



Mean Absolute Percentage Error (MAPE)
= 73.07256613839873%

MODEL COMPARISON

	OLS Regression	XGBoost	AdaBoost	Gradient Boosting
RMSE	0.0475	0.0509	0.0516	0.0478
MAPE (%)	73.2	78.9	105.8	73.0
Best Fit	✓			

OUR TAKEAWAYS



Using **techniques of handling** the data set, exploratory analysis, data visualization efficiently



Lower R² model does **not** mean that it is a **bad fit** for the dataset



One-hot encoding to **normalise** the vast range of the dataset



Utilise **powerful regression models** such as XGBoost, ADABoost and Gradient boosting



VIF Score & Backward Elimination



Version control using GitHub

CONCLUSION



Low Predictor & Response Correlation

- Drop "Grade 10 percentage"
- Highest Correlation: 0.24
- R^2 of our model will be low



Presence of Several Specialisations

Group those < 10 under "Others"



Normalised All Variables with One-Hot Encoding

Normalise the data and thus their influence on the results



Model Comparison

Best Fit Model - RSME & MAPE

CONCLUSION

Use

Ordinary Least Squares
Regression

for predicting

Engineering Graduates'
Salaries



**THANK
YOU!**

CONTENTS OF THIS TEMPLATE

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

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

TABLE OF CONTENTS

01

PROBLEM
DEFINITION

You can describe
the topic of the
section here

02

PROBLEM
DEFINITION

You can describe
the topic of the
section here

03

THE PRODUCT
WE HAVE

You can describe
the topic of the
section here

04

BUSINESS
MODEL

You can describe
the topic of the
section here

OUR TEAM



KALIYAH HARRIS

You can speak a bit about
this person here



PETER HILL

You can speak a bit about
this person here

WHOA!

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

INTRODUCTION

You can give a **brief description** of the topic you want to talk about here. If you want to talk about Mercury, you can say that it's the smallest planet in the Solar System

01

PROBLEM DEFINITION

You can enter a subtitle
here if you need it

CONCLUSION

MONTH 1

Venus is the second planet from the Sun



MONTH 2

Jupiter is the biggest planet

MONTH 3

Saturn is a gas giant with several rings



MONTH 4

Despite being red, Mars is cold



EDA3: COLLEGE TIER & CITY TIER (?)

MARS | 26%

Despite being red,
Mars is a cold place

VENUS | 17%

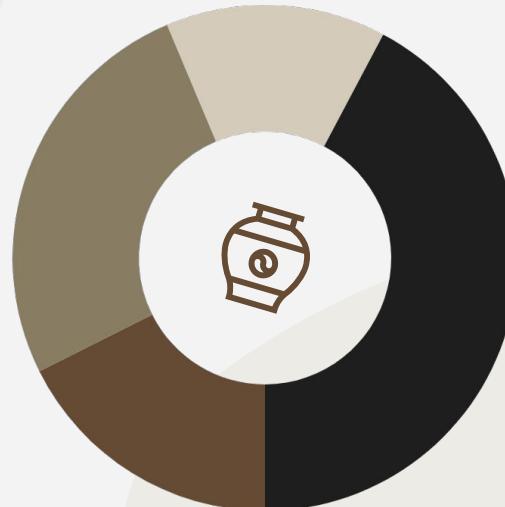
Venus the second
planet from the Sun

JUPITER | 14%

Jupiter is the
biggest planet

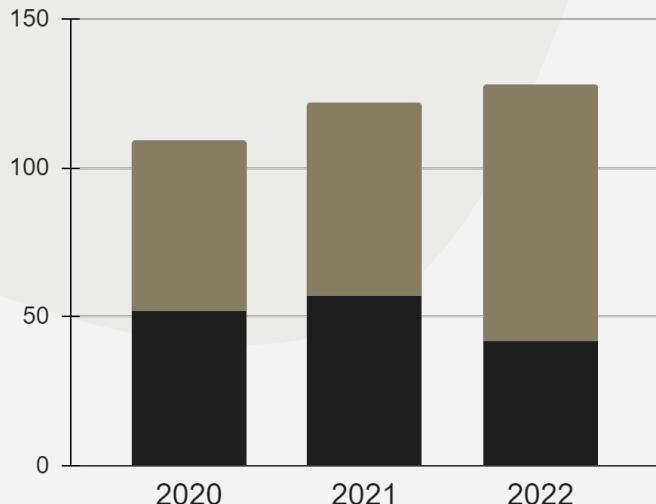
SATURN | 43%

Saturn is a gas giant
with several rings

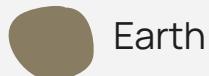


Follow the link in the graph to modify its data and then
paste the new one here. [For more info, click here](#)

EDA6: GENDER (?)



Mars



Earth



MERCURY

It's the closest planet to the Sun



NEPTUNE

It's the farthest planet from the Sun



JUPITER

It's the biggest planet of them all

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

A TIMELINE ALWAYS WORKS FINE

1996



2010



2016



2021



MERCURY

Mercury is the closest planet to the Sun

MARS

Despite being red, Mars is actually a cold place

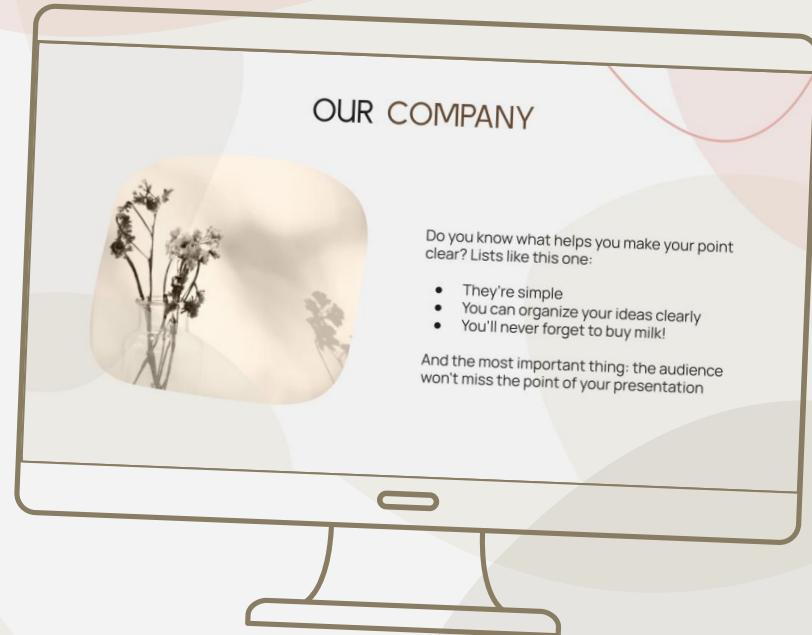
JUPITER

Jupiter is the biggest planet of them all

- Remove duplicates
- Remove null values
- Balance review scores

PRODUCT DEMO

You can **replace the image** on the screen with your own work. Right-click on it and then choose "Replace image" so you can add yours



A PICTURE ALWAYS REINFORCES THE CONCEPT

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



SWOT ANALYSIS

MARS

Despite being red, Mars is actually cold

JUPITER

It's the biggest planet in the Solar System

SATURN

Saturn is a gas giant and has several rings

MERCURY

Mercury is the closest planet to the Sun

S

W

O

T

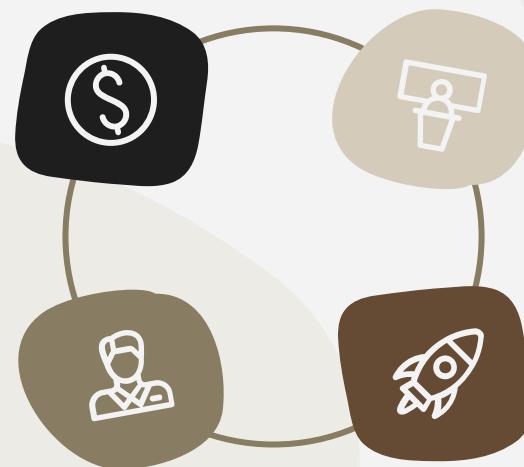
PRODUCT OVERVIEW

MARS

Mars is actually a very cold place

JUPITER

It's the biggest planet in the Solar System



NEPTUNE

It's the farthest planet from the Sun

SATURN

Saturn is a gas giant and has rings

OUR PLANS

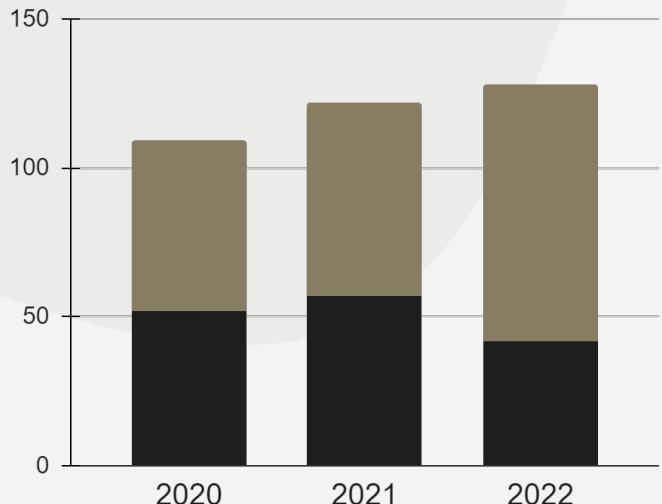
BASIC	PRO	PREMIUM	FREEMIUM
Mars is actually a very cold place	Jupiter is the biggest planet	Venus has a beautiful name	Saturn is a gas giant with rings
			
\$10	\$20	\$45	\$30

“

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

—SOMEONE FAMOUS

OUR TRACTION



Mars



Earth



MERCURY

It's the closest planet to the Sun



NEPTUNE

It's the farthest planet from the Sun



JUPITER

It's the biggest planet of them all

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

CASE STUDY



CHALLENGE

RESULT

SOLUTION

1ST PROJECT	Mercury	✓	Mars
2ND PROJECT	Neptune	✗	Saturn
3RD PROJECT	Earth	✓	Neptune
4ST PROJECT	Moon	✗	Venus



REVIEWS

EUN
YEONG

“Mercury is the closest planet to the Sun and the smallest one”

GYEONG
HUI

“Venus has a beautiful name and is the second from the Sun”

THIS IS A MAP



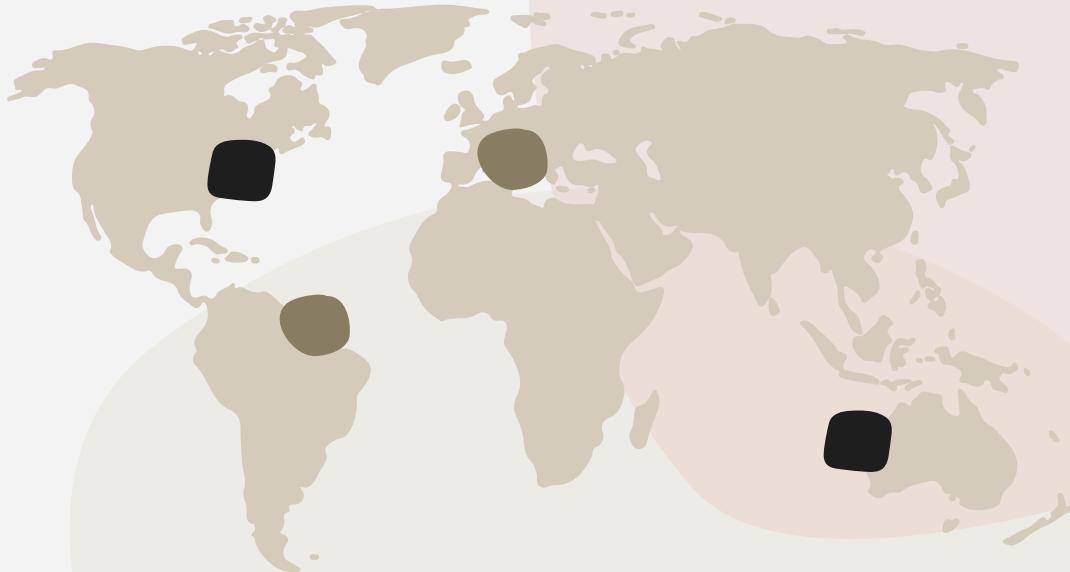
VENUS

It has a beautiful name, but it's hot



MERCURY

It's the closest planet to the Sun



AWARDS



MARS

Mars is actually a very cold place



NEPTUNE

It's the farthest planet from the Sun



MERCURY

It's the closest planet to the Sun



SATURN

Saturn is the ringed planet and a gas giant



JUPITER

Jupiter is the biggest planet of them all



VENUS

Venus is the second planet from the Sun



AWESOME WORDS

MARKET SIZE

MARS | 26%

Despite being red,
Mars is a cold place

VENUS | 17%

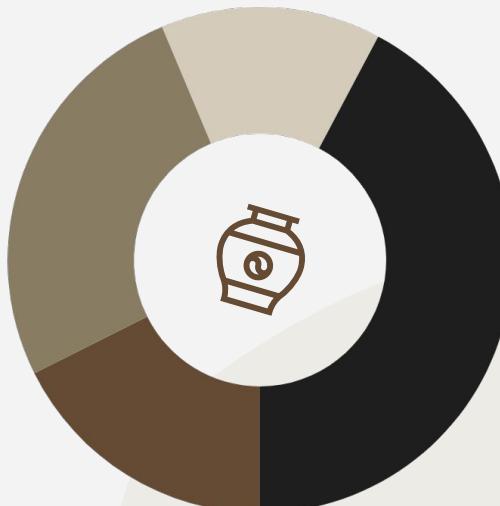
Venus the second
planet from the Sun

JUPITER | 14%

Jupiter is the
biggest planet

SATURN | 43%

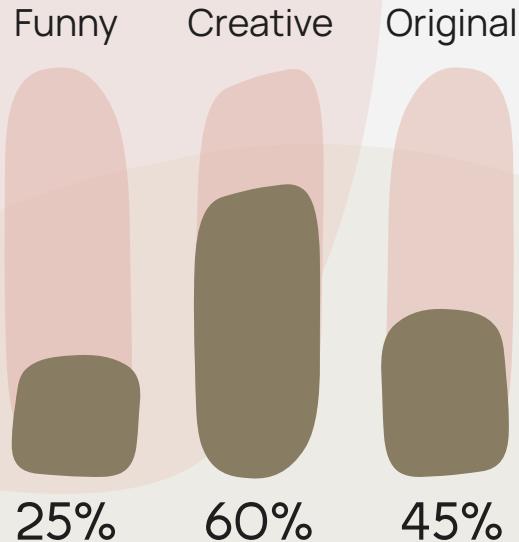
Saturn is a gas giant
with several rings



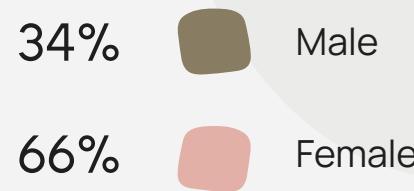
Follow the link in the graph to modify its data and then
paste the new one here. [For more info, click here](#)

OUR TARGET

PERSONALITY



GENDER



HOUSEHOLD INCOME



COMPETITORS



MARS

It's actually a cold place



NEPTUNE

Neptune is far away from Earth



JUPITER

It's the biggest planet



MERCURY

It's the smallest planet



SATURN

Saturn is the ringed planet



VENUS

Venus has a beautiful name

OUR TIMING

MONTH 1

Venus is the second planet from the Sun



MONTH 2

Jupiter is the biggest planet

MONTH 3

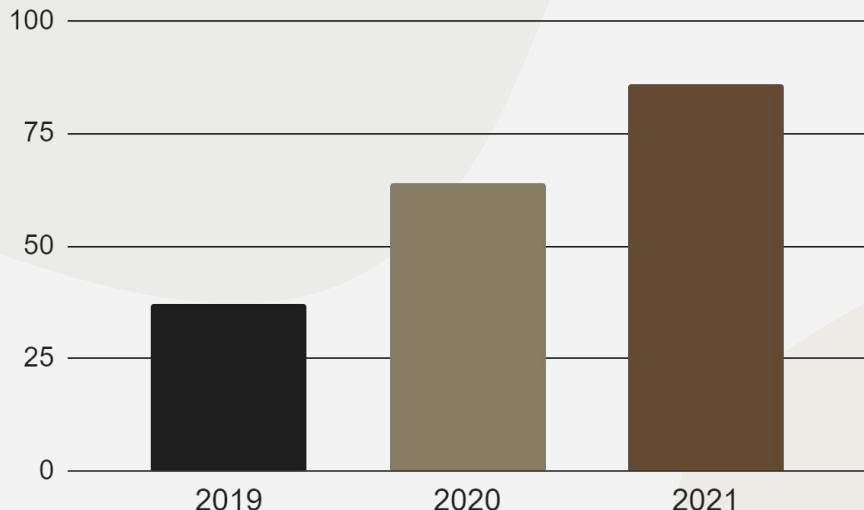
Saturn is a gas giant with several rings



MONTH 4

Despite being red, Mars is cold

PREDICTED GROWTH



MERCURY

It's the smallest planet



SATURN

Saturn is the ringed planet



VENUS

Venus has a beautiful name

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

INVESTMENT



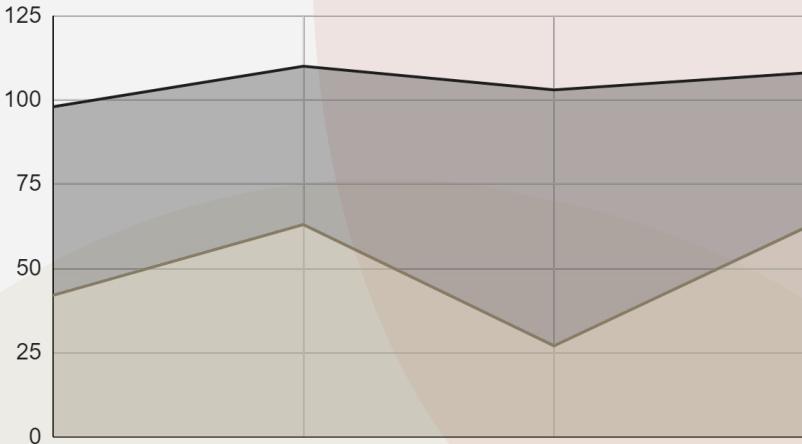
VENUS

Venus has a beautiful name



MERCURY

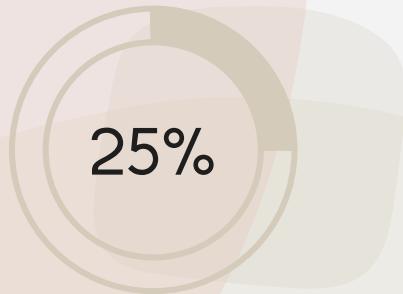
Mercury is the smallest planet



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

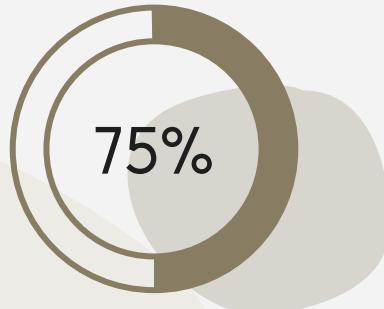
MONTHLY ANALYSIS

MERCURY



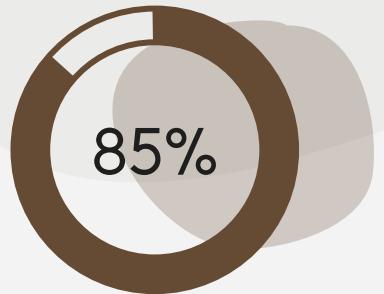
Mercury is the closest planet to the Sun and the smallest one

VENUS

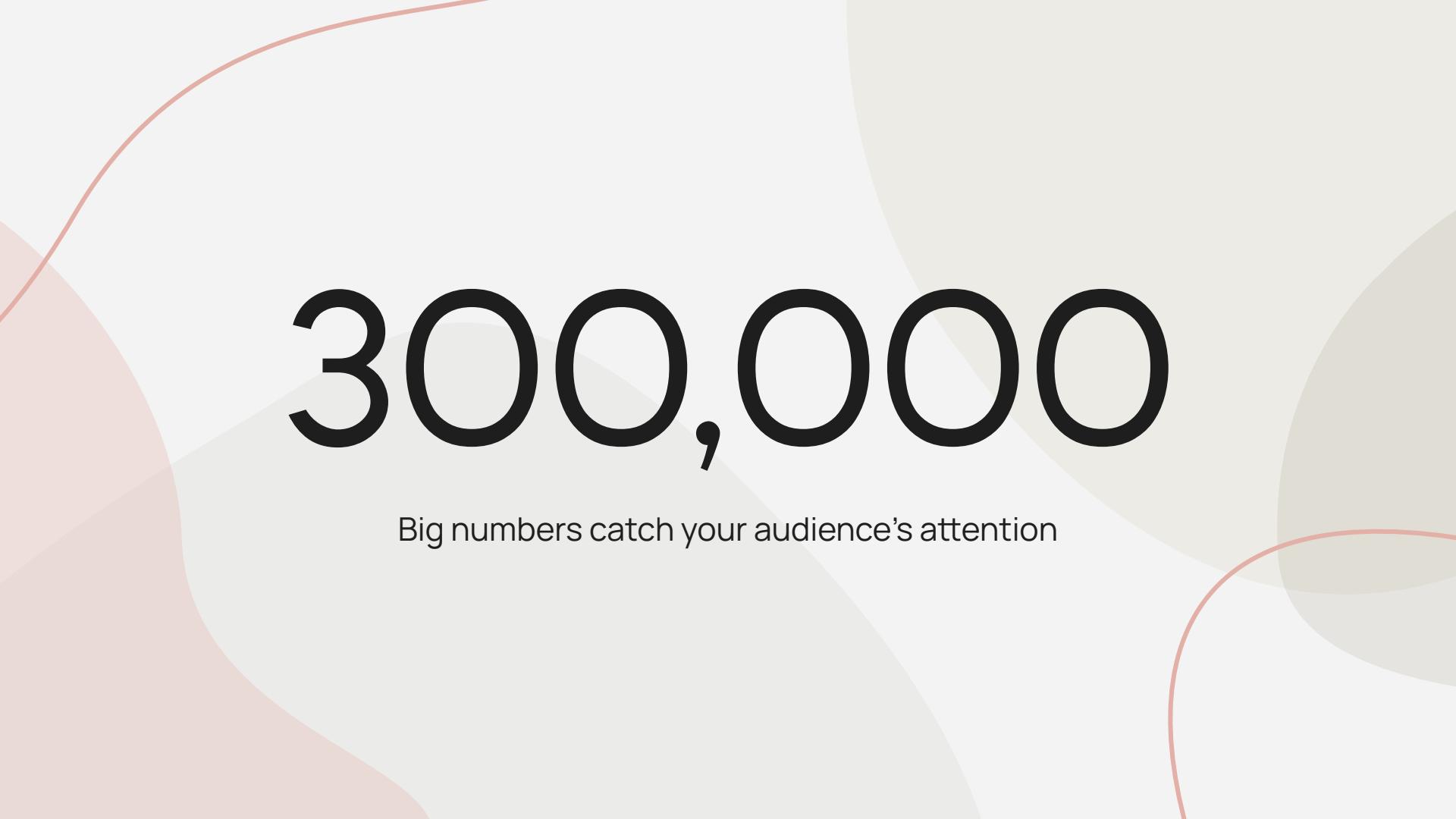


Venus has a beautiful name and is the second planet from the Sun

MARS



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



300,000

Big numbers catch your audience's attention

9h 55m

Jupiter's rotation period

333,000

The Sun's mass compared to Earth's

A PICTURE
IS WORTH A
THOUSAND
WORDS



A TIMELINE ALWAYS WORKS FINE

1996



VENUS

Venus is the second planet from the Sun

2010



MERCURY

Mercury is the closest planet to the Sun

2016



MARS

Despite being red, Mars is actually a cold place

2021



JUPITER

Jupiter is the biggest planet of them all

THANKS!

Does anyone have any questions?

addyouremail@freepik.com

+91 620 421 838

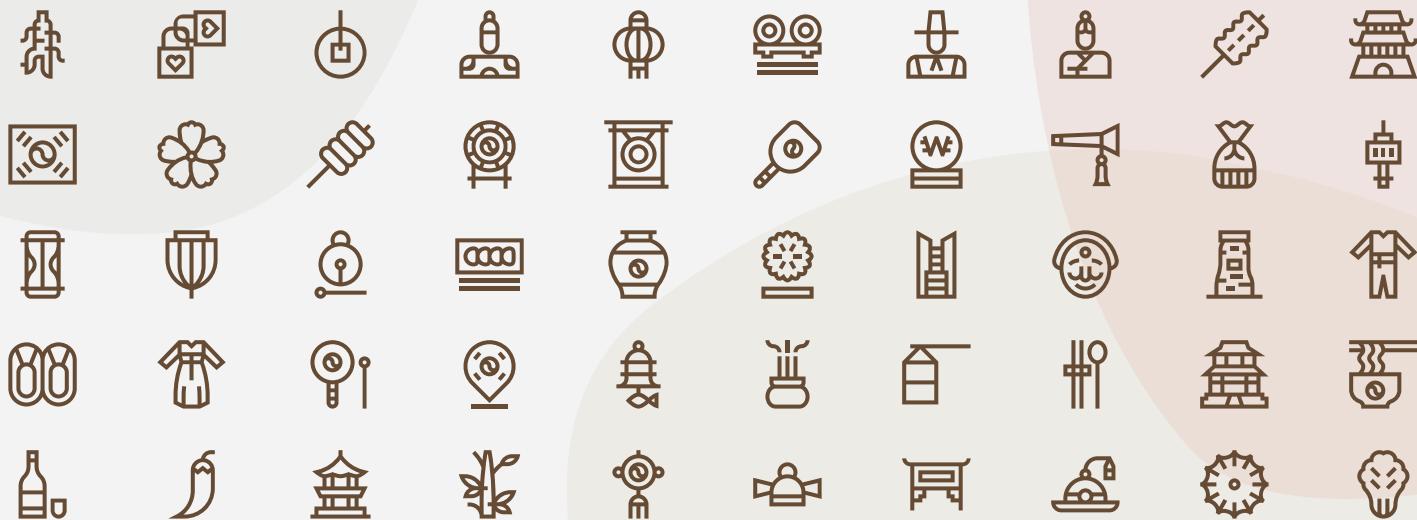
yourcompany.com



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

Please keep this slide for attribution

ICON PACK: SEOUL | LINEAL



ALTERNATIVE RESOURCES

Here's an assortment of alternative resources whose style fits the one of this template:



RESOURCES

ICONS

- Icon Pack: Seoul | Lineal

PHOTOS

- Abstract minimal concept flowers and shadows
 - People wearing k-pop aesthetics clothing
- Abstract minimal concept flowers and shadows
 - Friends wearing k-pop aesthetics clothing 1
 - White still life composition

VECTORS

- Creative pastel painted wallpaper

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)

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:

Kulim Park

(<https://fonts.google.com/specimen/Kulim+Park>)

Manrope

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

#1e1e1e

#664b34

#887c62

#d4cbbb

#e7e2d6

#f3f3f3

#e2b0a6

#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



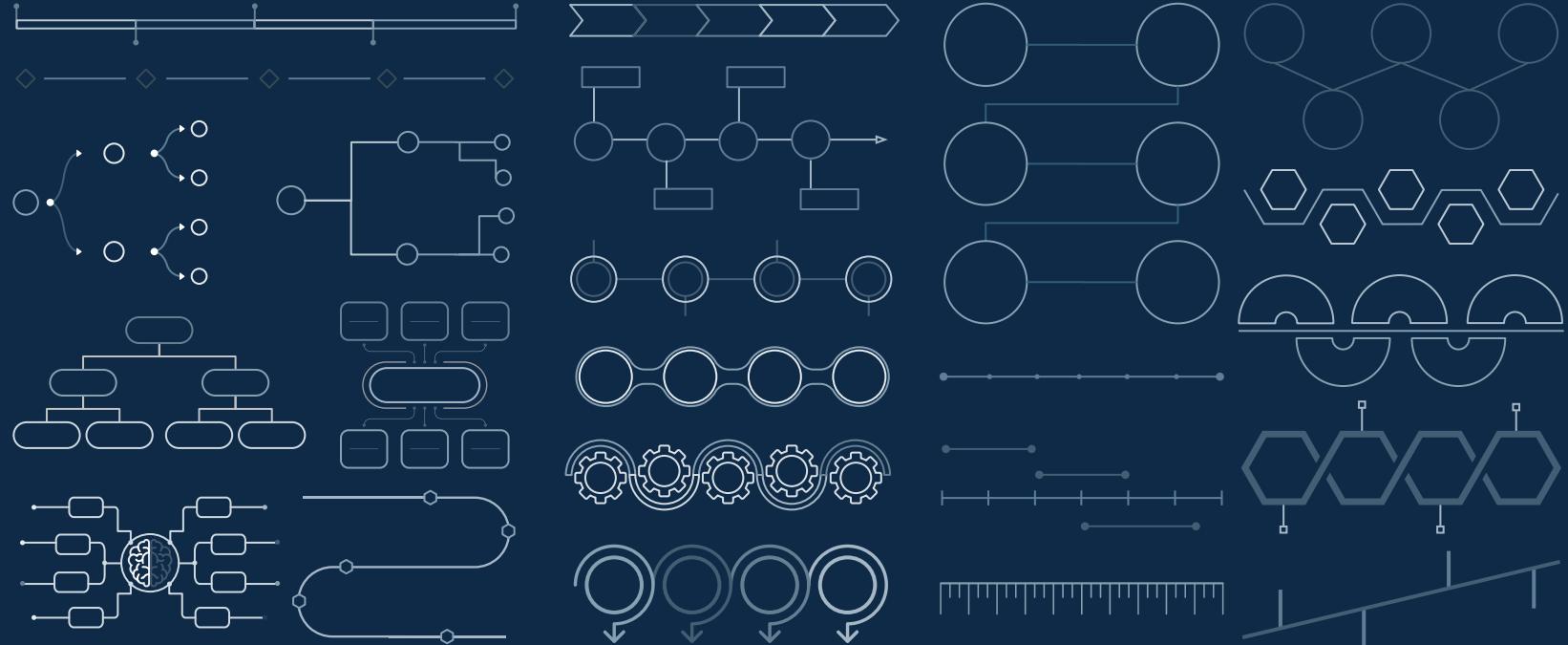
Cuate

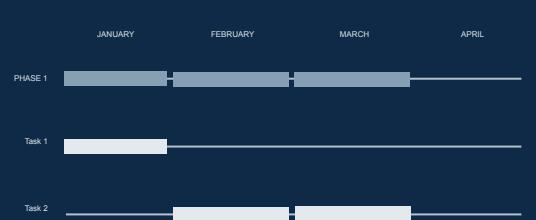
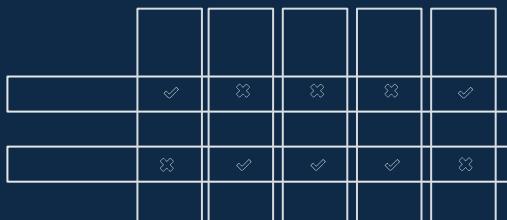
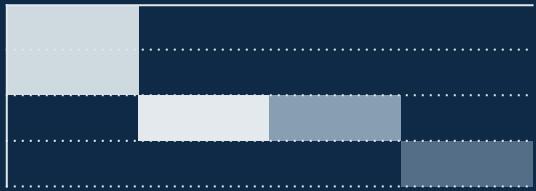
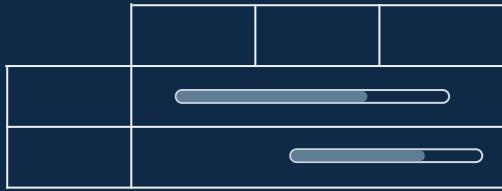
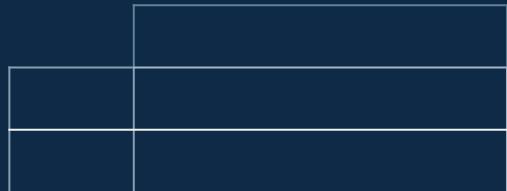
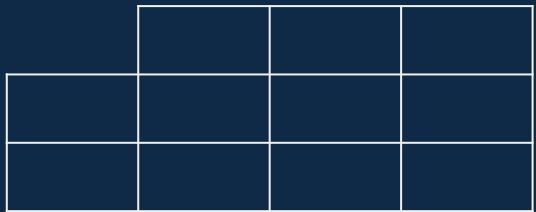
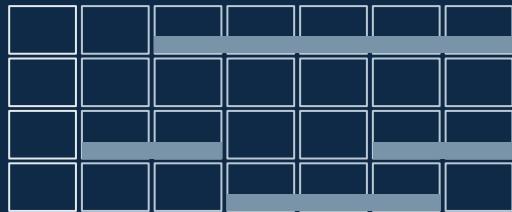
Use our editable graphic resources...

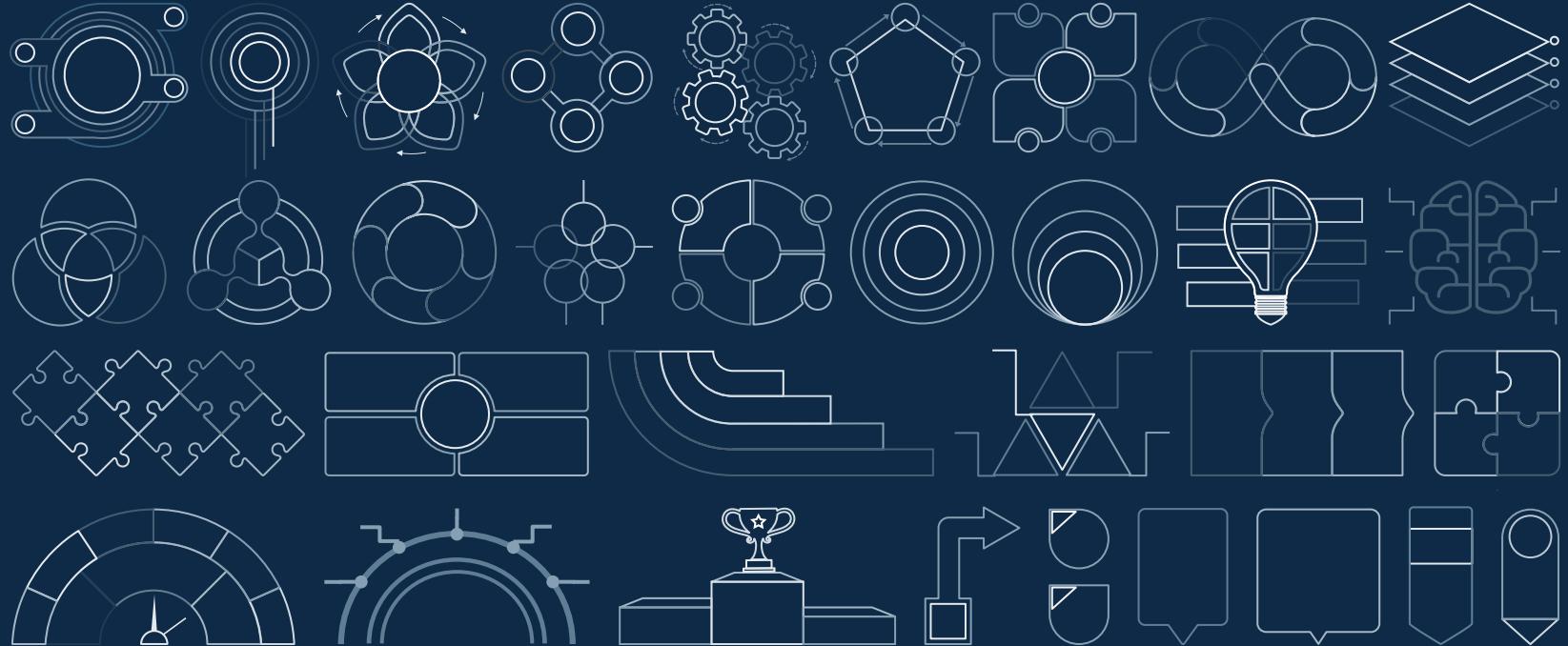
You can easily resize these resources without losing quality. To change the color, just ungroup the resource and click on the object you want to change. Then, click on the paint bucket and select the color you want. Group the resource again when you're done. You can also look for more infographics on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

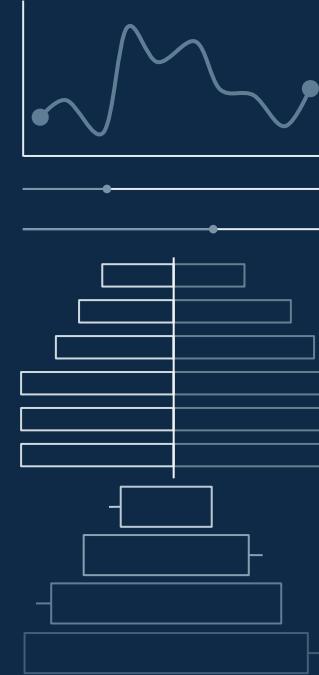
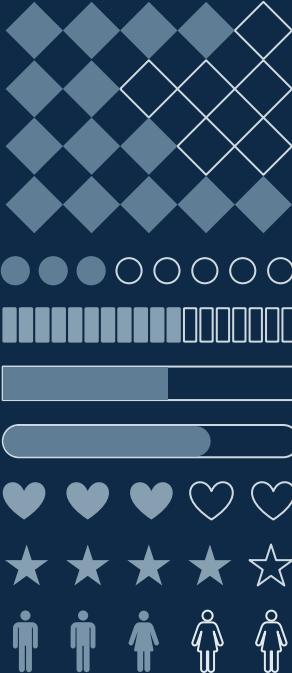
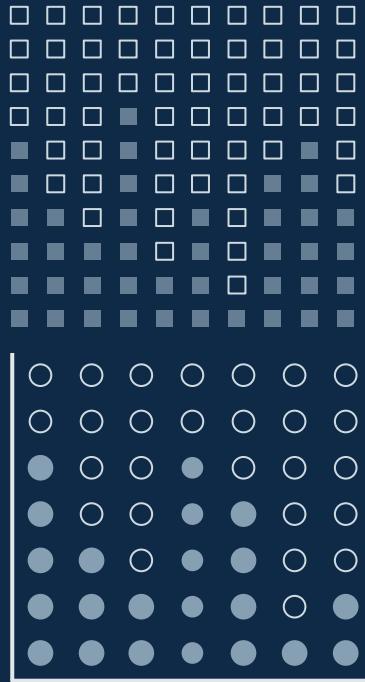












...and our sets of editable icons

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



