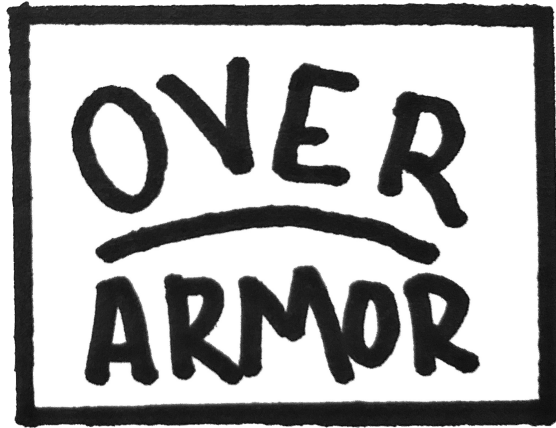


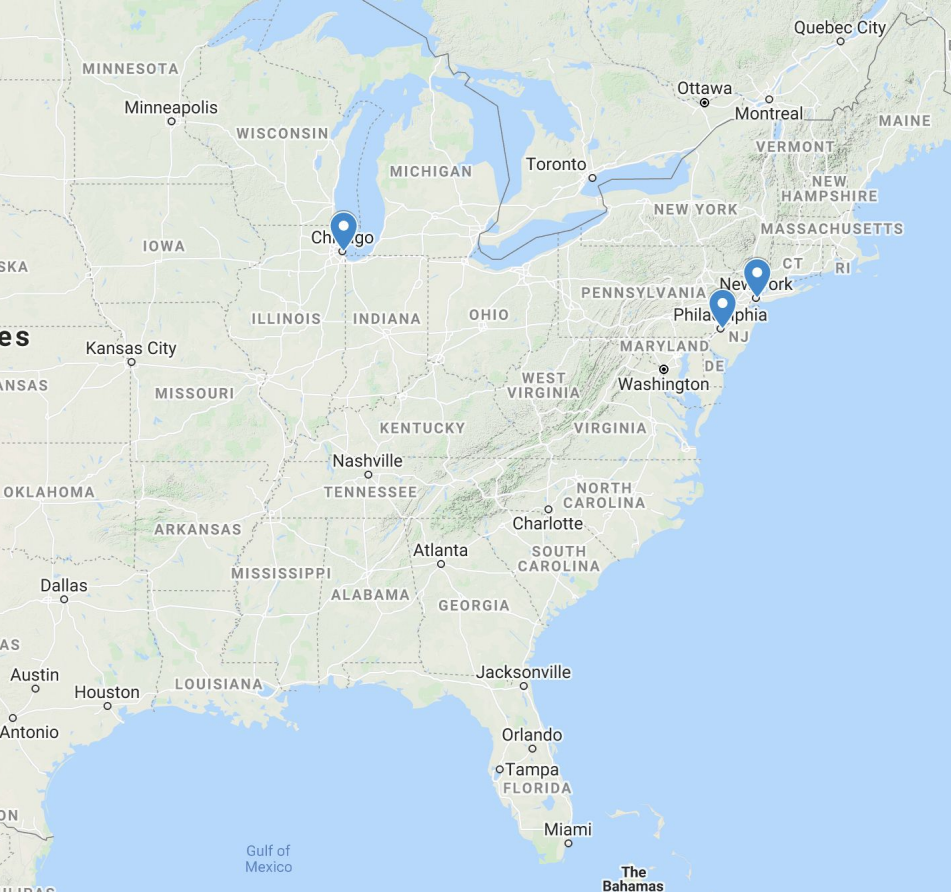
Market Research: Sports Fans in the U.S. and England

Riley Robertson

The Client



The Client



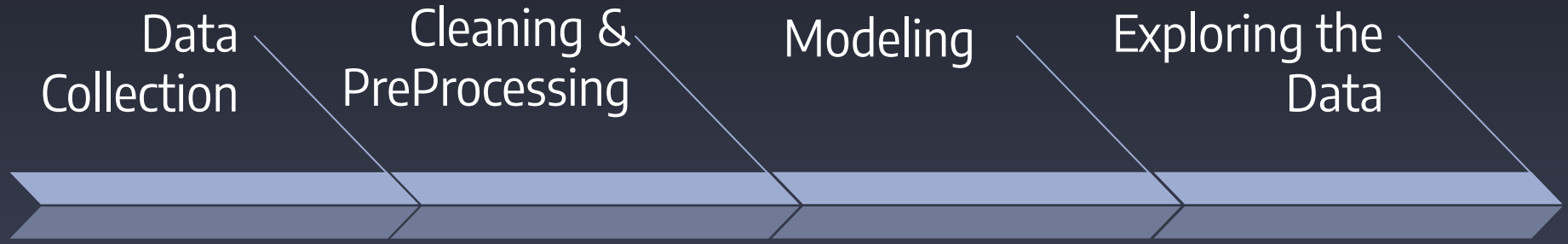
The Client



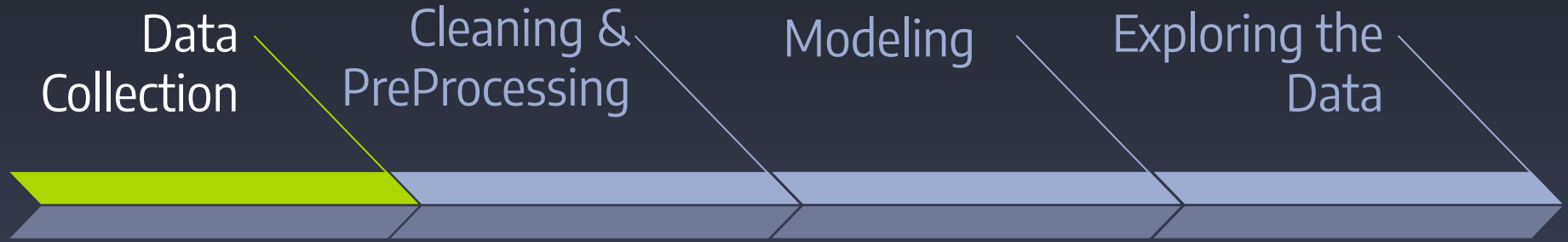
OverArmor's Requests:

1. Collect data from each league's respective Reddit community to build a clean database for investigating the colloquial language used by each fan-base
2. Based on that data, create a machine learning model that can accurately classify a post if the source is unknown
3. Provide insight into the vernacular of each fanbase by creating lists of the most used and most significant words to each community

Workflow Overview

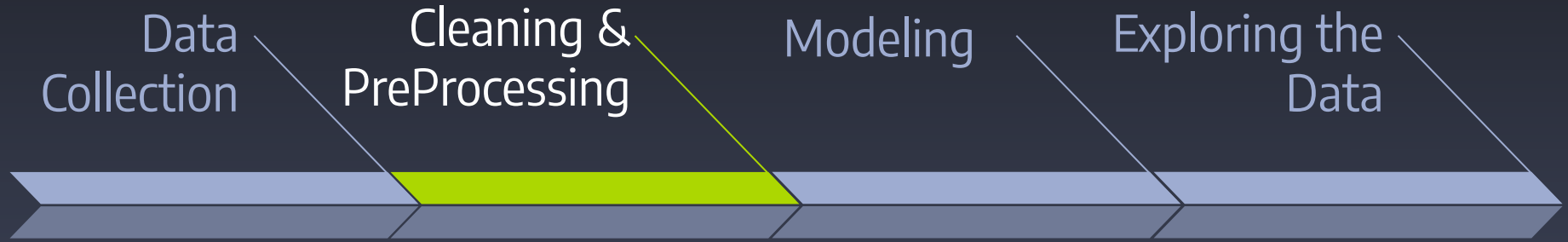


Workflow Overview



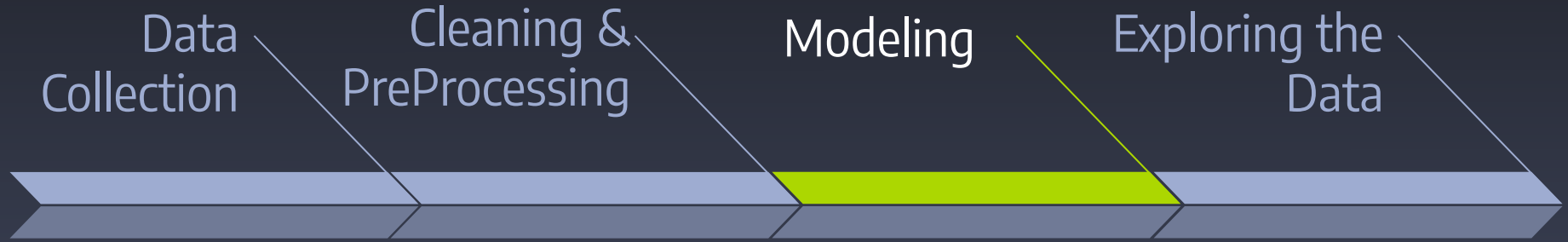
- ▷ 100k posts per subreddit
- ▷ Lots of cleaning required, but room for refinement and balancing the two subs

Workflow Overview



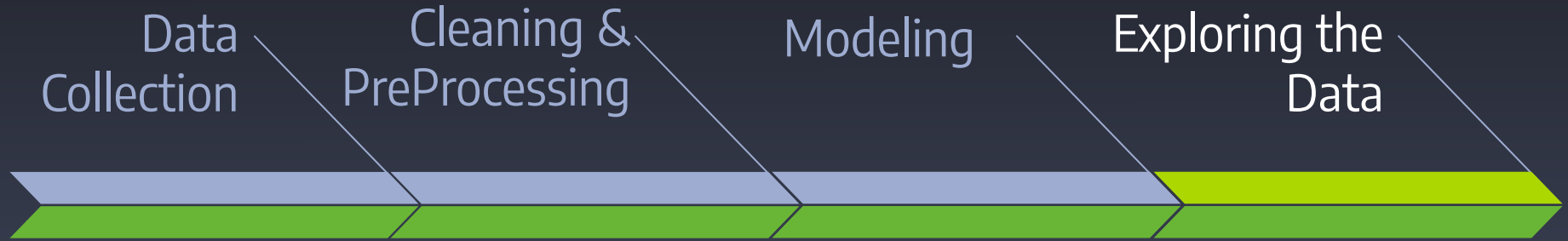
- ▷ Decreased to 12k total posts with a 56% / 44% split
- ▷ That number serves as my baseline - the number to beat

Workflow Overview



- Logistic Regression and Random Forest both returned excellent results

Workflow Overview

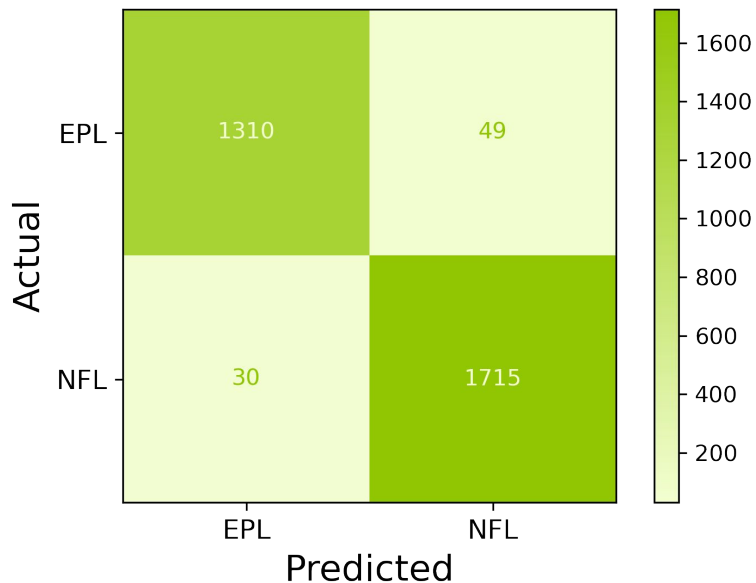


- ▷ EDA produced several lists I can provide to OverArmor that will help them craft their messaging.

Model Comparison

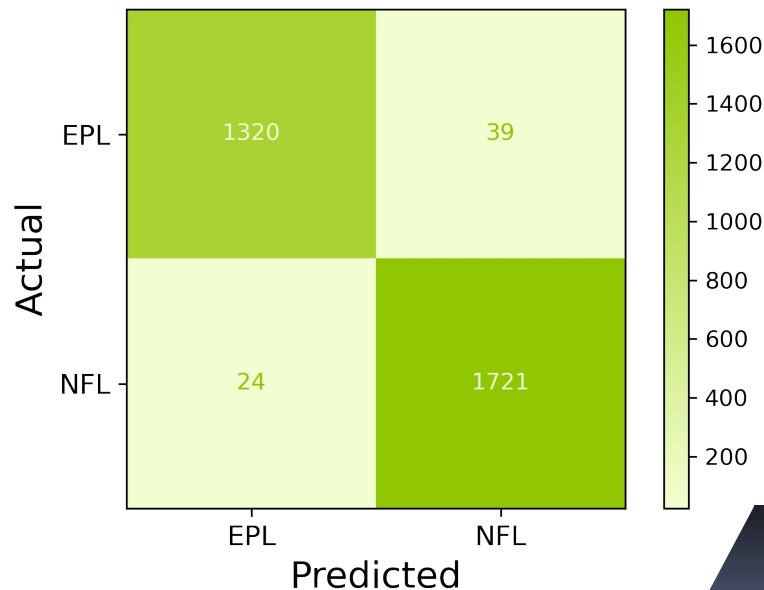
Logistic Regression

97.45% Accuracy



Random Forest

97.71% Accuracy



The Final Request

	count_all	freq%_all	count_nfl	freq%_nfl	count_epl	freq%_epl
team	7688.0	0.6214031684448755	5510.0	0.7892852026930239	2178.0	0.4040066777963272
season	5566.0	0.4498868412544455	3628.0	0.5196963185790001	1938.0	0.35948803561491377
game	5391.0	0.4357419980601358	4063.0	0.5820083082652915	1328.0	0.24633648673715453
league	5150.0	0.4162625282896864	1362.0	0.1951009883970778	3788.0	0.7026525690966425
like	5049.0	0.40809893307468476	3428.0	0.4910471279186363	1621.0	0.30068632906696346
nfl	4945.0	0.399692854833495	4898.0	0.7016186792723106		
just	4720.0	0.3815066278693825	3175.0	0.454805901733276	1545.0	0.2865887590428492
teams	4360.0	0.35240866472680243	3100.0	0.4440624552356396	1260.0	0.2337228714524207
think	4195.0	0.3390720982864533	2805.0	0.40180489901160293	1390.0	0.25783713596735297
year	4026.0	0.3254122211445199	3450.0	0.4941985388912763	576.0	0.10684474123539232
players	3430.0	0.2772389266084707	2304.0	0.3300386764073915	1126.0	0.2088666295677982
games	3389.0	0.27392499191723246	2401.0	0.34393353367766795	988.0	0.18326841031348545
time	3187.0	0.2575978014872292	2308.0	0.3306116602205988	879.0	0.16304952698942682
football	3172.0	0.2563853863562884	1817.0	0.2602778971494055	1355.0	0.2513448339825635
know	2721.0	0.2199321047526673	1716.0	0.24581005586592178	1005.0	0.1864218141346689
best	2605.0	0.21055609440672488	1694.0	0.24265864489328176	911.0	0.1689853459469486
im	2567.0	0.20748464274167475	1675.0	0.2399369717805472	892.0	0.16546095344092004
play	2542.0	0.20546395085677335	1910.0	0.27359977080647474	632.0	0.11723242441105547
qb	2513.0	0.20311994827028776	2513.0	0.3599770806474717		
good	2510.0	0.20287746524409958	1775.0	0.2542615671107291	735.0	0.13633834168057873
player	2427.0	0.19616876818622697	1708.0	0.24466408823950722	719.0	0.13337043220181785
years	2338.0	0.188975105075978	1818.0	0.26042114310270736	520.0	0.096457058089752918
week	2248.0	0.1817006142903302	1917.0	0.27460249247958746	331.0	0.061398627341866074
premier	2241.0	0.1811348205625606			2220.0	0.41179744017807457
10	2238.0	0.18089233753637246	1830.0	0.26214009454232917	408.0	0.07568169170840289
new	2076.0	0.16779825412221144	1421.0	0.20355249964188513	655.0	0.12149879428677425
super	2039.0	0.1648076301325574	1857.0	0.2660077352814783	182.0	0.03375997032090521
really	1946.0	0.15729065632072423	1302.0	0.18650623119896864	644.0	0.11945835652012614
win	1940.0	0.15680569026834787	1379.0	0.19753616960320872	561.0	0.10406232609905398
bowl	1894.0	0.1530876172001293	1890.0	0.2707348517404383		
yards	1891.0	0.15284513417394116	1880.0	0.26930239220742014		

Subreddit	Rank	Token	Category	Odds
EPL	1	Discussion	Tag	9.527237
EPL	2	premier	League Name	5.964373
EPL	3	Question	Tag	5.710886
EPL	4	liverpool	Team Name	2.853200
EPL	5	pl	League Name	2.770702
EPL	6	club	Vocab	2.330990
EPL	7	league	Vocab	2.291374
EPL	8	[flair] Premier League	Tag	2.274725
EPL	9	arsenal	Team Name	2.154184
EPL	10	epl	League Name	2.131302
EPL	11	chelsea	Team Name	2.081810
EPL	12	var	Vocab - Acc	1.841295
EPL	13	united	Team Name	1.702915
EPL	14	match	Vocab	1.672153
EPL	15	poll	Error	1.655469
EPL	16	spurs	Team Name	1.641890
EPL	17	tottenham	Team Name	1.629748
EPL	18	city	Team Name	1.571513
EPL	19	english	League Name	1.532463
EPL	20	News	Tag	1.469784
EPL	21	clubs	Vocab	1.449543
EPL	22	everton	Team Name	1.441820
EPL	23	goals	Vocab	1.428311
EPL	24	[flair] Manchester United	Tag	1.406905

Column Name	Feature Importance
nfl	0.08079127506331457
premier	0.04195555008195937
league	0.031593330792543534
qb	0.029927934064901993
liverpool	0.019435090819427892
year	0.019291671958764037
draft	0.01608605003842847
united	0.015534190807037546
bowl	0.015438240773194035
yards	0.012896255309559327
club	0.01283742908792805
chelsea	0.012389190595917756
brady	0.012012156557565346
arsenal	0.011464017687293389
team	0.011151358319933918

Conclusions

1. Collected and cleaned data
2. Created 2 models that prioritized accuracy and succeeded in classifying posts at 97+% rates
3. Gathered various lists of language used in each subreddit

With more time:

1. Dive much deeper into the language
2. Investigate misidentified posts
3. Run further testing on posts my models haven't seen





The Client





Arguably the most popular sports leagues in each of their respective countries, the National Football League (NFL) and the English Premier League (EPL) boast some of the largest fan-bases in North America and Europe, making them prime targets for advertising.



A picture is worth a thousand words

A complex idea can be conveyed with just a single still image, namely making it possible to absorb large amounts of data quickly.





Want big impact? Use big image.

And tables to compare data

	A	B	C
Yellow	10	20	7
Blue	30	15	10
Orange	5	24	16

Data
Collection

Cleaning

20XX

20XX

**100k Posts
Per Subreddit**

**Decreased to
13k total with
a 56/44 Split**

Lorem Ipsum

Lorem ipsum dolor sit amet, consectetur
adipiscing.

Lorem Ipsum

Lorem ipsum dolor sit amet, consectetur
adipiscing.

Maps



89,526,124

Whoa! That's a big number, aren't you proud?

15 million

NFL average viewers per game

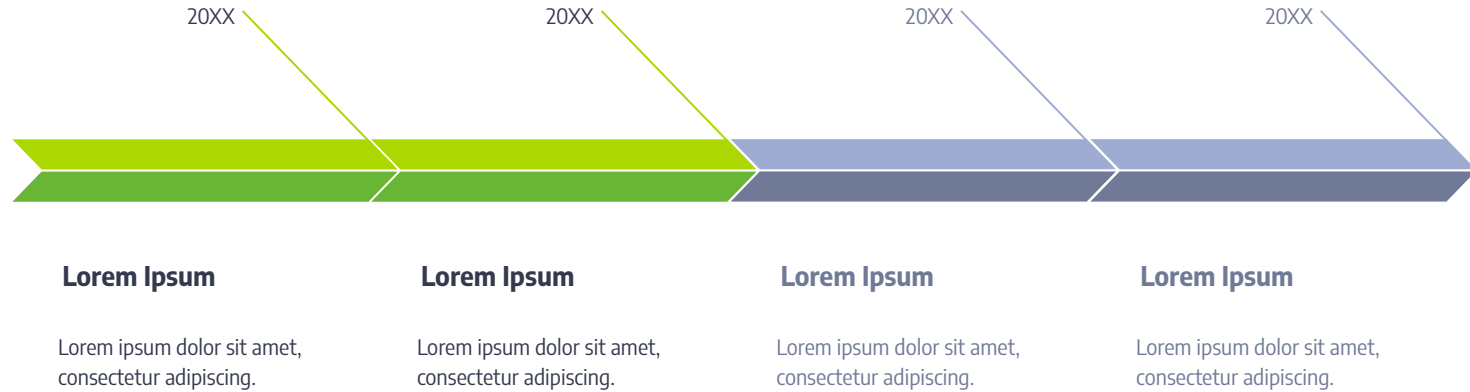
185,244 users

And a lot of users

100%

Total success!

Our process is easy



Let's review some concepts

Yellow

Is the color of gold, butter and ripe lemons. In the spectrum of visible light, yellow is found between green and orange.

Yellow

Is the color of gold, butter and ripe lemons. In the spectrum of visible light, yellow is found between green and orange.

Blue

Is the colour of the clear sky and the deep sea. It is located between violet and green on the optical spectrum.

Blue

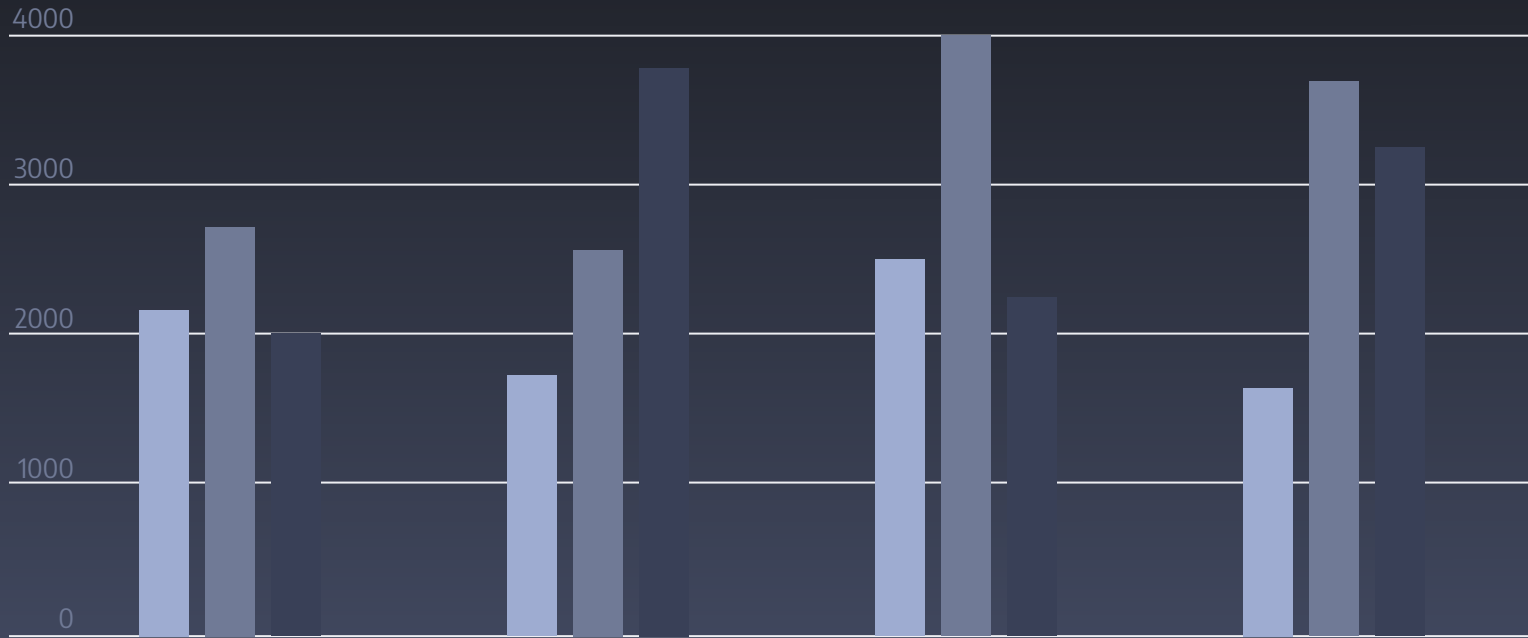
Is the colour of the clear sky and the deep sea. It is located between violet and green on the optical spectrum.

Red

Is the color of blood, and because of this it has historically been associated with sacrifice, danger and courage.

Red

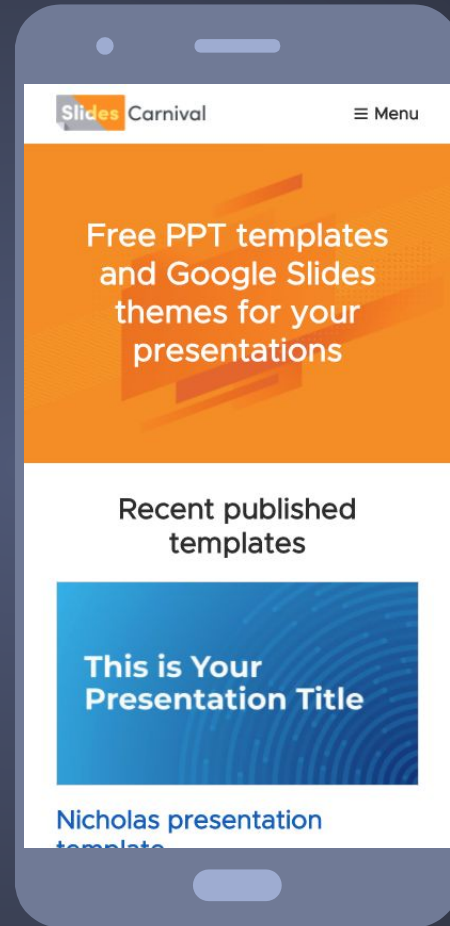
Is the color of blood, and because of this it has historically been associated with sacrifice, danger and courage.



You can insert graphs from Excel or Google Sheets

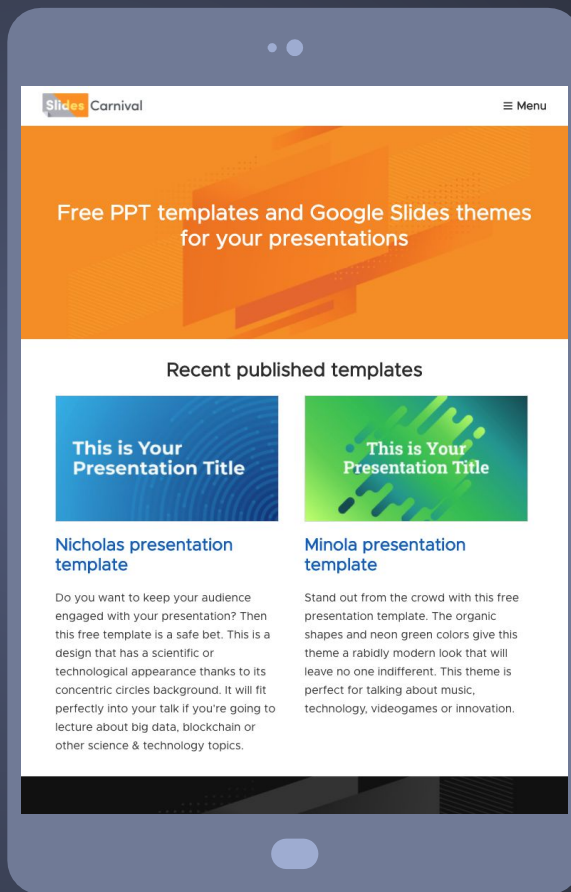
Mobile project

Show and explain your web, app or software projects using these gadget templates.



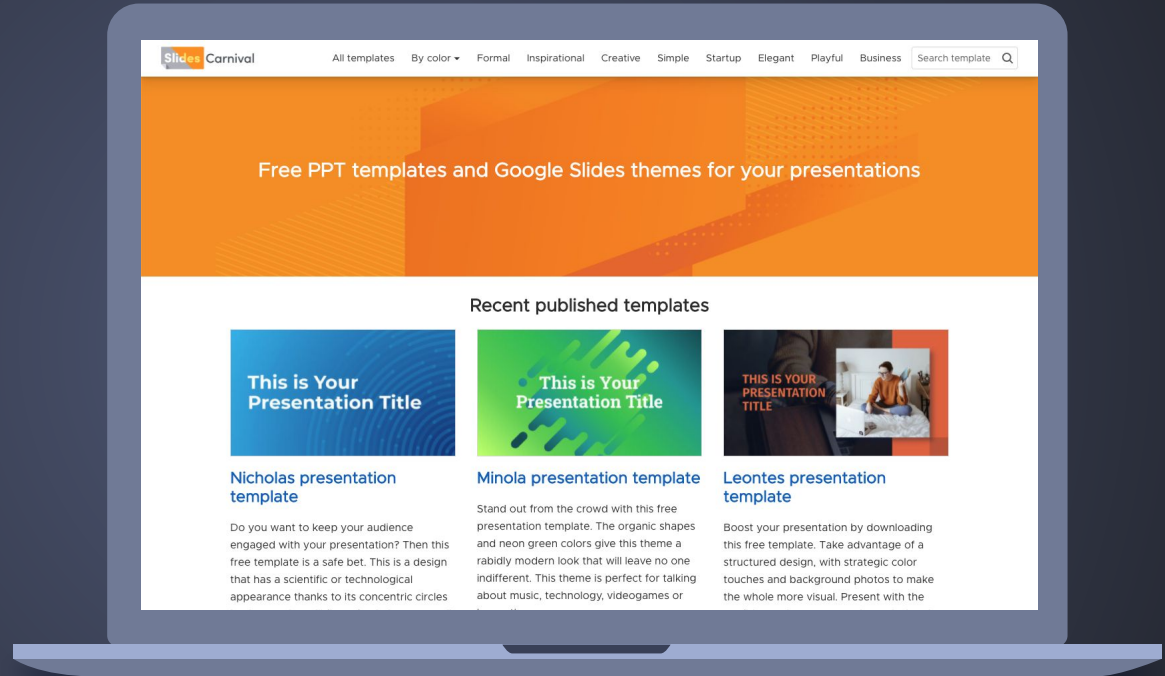
Tablet project

Show and explain your web, app or software projects using these gadget templates.



Desktop project

Show and explain your web, app or software projects using these gadget templates.



Thanks!

Any questions?

You can find me at @username and user@mail.me

Credits

Special thanks to all the people who made and released these awesome resources for free:

- ▷ Presentation template by [SlidesCarnival](#)
- ▷ Photographs by [Unsplash](#)

Presentation design

This presentation uses the following typographies:

- ▷ Titles: Encode Sans Semi Condensed
- ▷ Body copy: Encode Sans Semi Condensed

Download for free at:

<https://www.fontsquirrel.com/fonts/encode-sans>

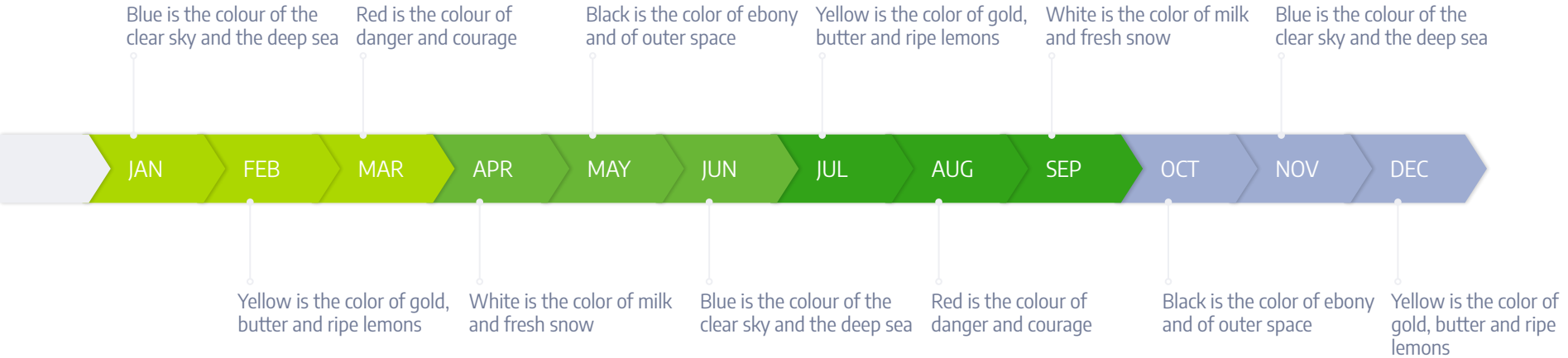
You don't need to keep this slide in your presentation. It's only here to serve you as a design guide if you need to create new slides or download the fonts to edit the presentation in PowerPoint®

2

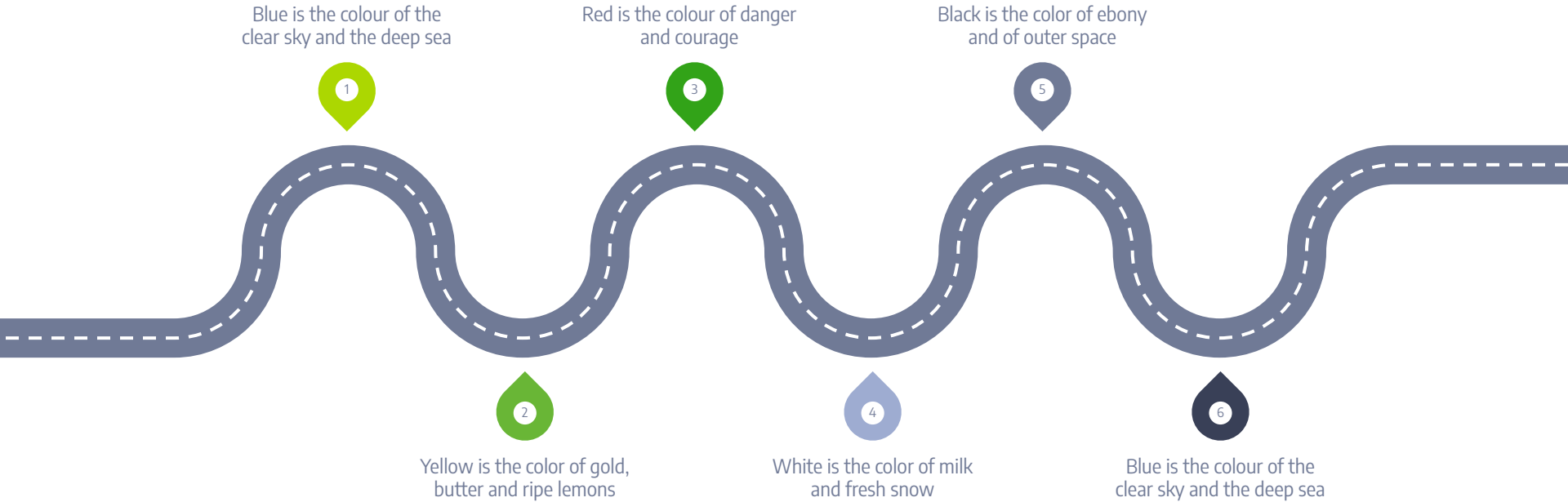
Extra Resources

For Business Plans, Marketing Plans, Project Proposals, Lessons, etc

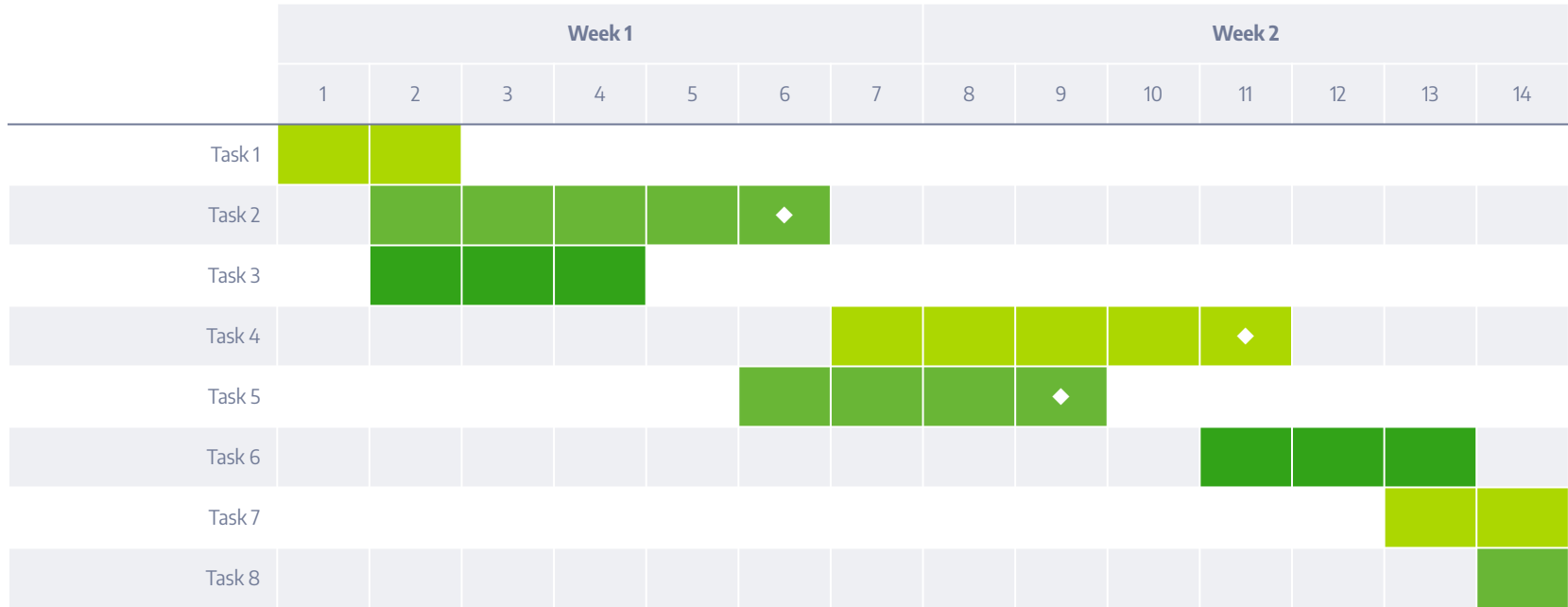
Timeline



Roadmap



Gantt chart



Diagrams and infographics

