



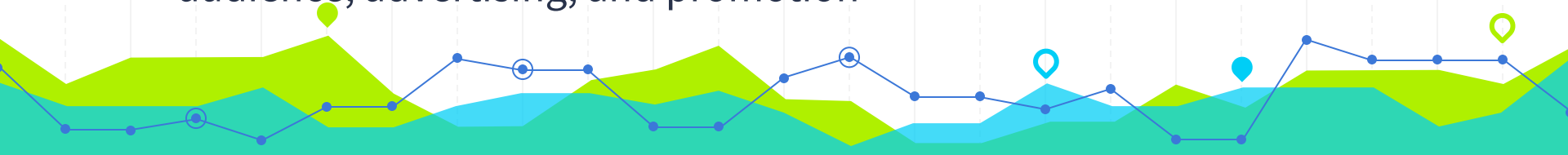
Bank Marketing

**By Wilbert Rodriguez and
Tommy Le**

Bank Marketing



- Bank Marketing refers to strategies that are undertaken by banks
 - Promote their products and services to customers
 - Attract new clients
- It helps build brand awareness, increases customer engagement, and ultimately helps the business growth of the bank
- Some of its components are product promotion, target audience, advertising, and promotion



The Bank Marketing Data

- The Data we obtained is on direct marketing campaigns(phone calls) of a Portuguese banking institution
- The classification goal is to predict if the client will subscribe a term deposit

	age	job	marital	education	default	balance	housing	loan	contact	day	month	duration	campaign	pdays	previous	poutcome	y
0	58	management	married	tertiary	no	2143	yes	no	unknown	5	may	261	1	-1	0	unknown	no
1	44	technician	single	secondary	no	29	yes	no	unknown	5	may	151	1	-1	0	unknown	no
2	33	entrepreneur	married	secondary	no	2	yes	yes	unknown	5	may	76	1	-1	0	unknown	no
3	47	blue-collar	married	unknown	no	1506	yes	no	unknown	5	may	92	1	-1	0	unknown	no
4	33	unknown	single	unknown	no	1	no	no	unknown	5	may	198	1	-1	0	unknown	no
...
45206	51	technician	married	tertiary	no	825	no	no	cellular	17	nov	977	3	-1	0	unknown	yes
45207	71	retired	divorced	primary	no	1729	no	no	cellular	17	nov	456	2	-1	0	unknown	yes
45208	72	retired	married	secondary	no	5715	no	no	cellular	17	nov	1127	5	184	3	success	yes
45209	57	blue-collar	married	secondary	no	668	no	no	telephone	17	nov	508	4	-1	0	unknown	no
45210	37	entrepreneur	married	secondary	no	2971	no	no	cellular	17	nov	361	2	188	11	other	no

Data Before and After Preprocessing

	age	job	marital	education	default	balance	housing	loan	contact	day	month	duration	campaign	pdays	previous	outcome	y
0	58	management	married	tertiary	no	2143	yes	no	unknown	5	may	261	1	-1	0	unknown	no
1	44	technician	single	secondary	no	29	yes	no	unknown	5	may	151	1	-1	0	unknown	no
2	33	entrepreneur	married	secondary	no	2	yes	yes	unknown	5	may	76	1	-1	0	unknown	no
3	47	blue-collar	married	unknown	no	1506	yes	no	unknown	5	may	92	1	-1	0	unknown	no
4	33	unknown	single	unknown	no	1	no	no	unknown	5	may	198	1	-1	0	unknown	no
...
45206	51	technician	married	tertiary	no	825	no	no	cellular	17	nov	977	3	-1	0	unknown	yes
45207	71	retired	divorced	primary	no	1729	no	no	cellular	17	nov	456	2	-1	0	unknown	yes
45208	72	retired	married	secondary	no	5715	no	no	cellular	17	nov	1127	5	184	3	success	yes
45209	57	blue-collar	married	secondary	no	668	no	no	telephone	17	nov	508	4	-1	0	unknown	no
45210	37	entrepreneur	married	secondary	no	2971	no	no	cellular	17	nov	361	2	188	11	other	no

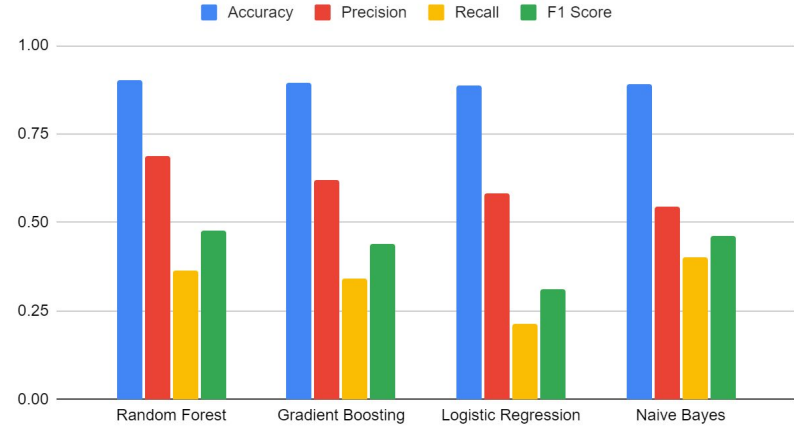
- The dataset didn't have any missing data
- It did have a lot of categorical features.
- This is where we used one-hot encoding

	age	balance	day	duration	campaign	pdays	previous	balance at age	job_admin.	job_blue- collar	...	career_technician_tertiary	career_technician_unknown	care
0	58	2143	5	261	1	-1	0	2201	0	0	...	0	0	
1	44	29	5	151	1	-1	0	73	0	0	...	0	0	
2	33	2	5	76	1	-1	0	35	0	0	...	0	0	
3	47	1506	5	92	1	-1	0	1553	0	1	...	0	0	
4	33	1	5	198	1	-1	0	34	0	0	...	0	0	
...
45206	51	825	17	977	3	-1	0	876	0	0	...	1	0	
45207	71	1729	17	456	2	-1	0	1800	0	0	...	0	0	
45208	72	5715	17	1127	5	184	3	5787	0	0	...	0	0	
45209	57	668	17	508	4	-1	0	725	0	1	...	0	0	
45210	37	2971	17	361	2	188	11	3008	0	0	...	0	0	

Machine Learning Techniques

- Random Forest
- Gradient Boosting
- Logistic Regression
- Naive Bayes

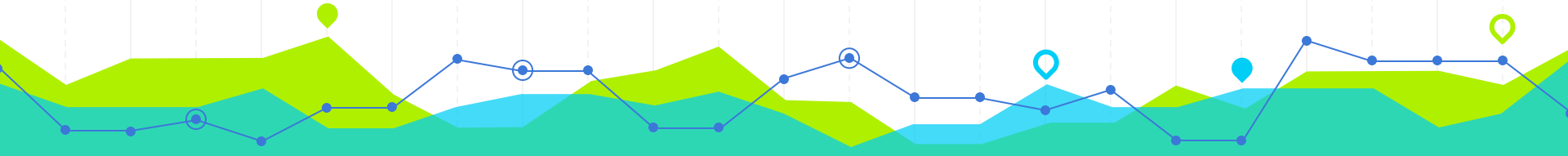
Accuracy, Precision, Recall and F1 Score

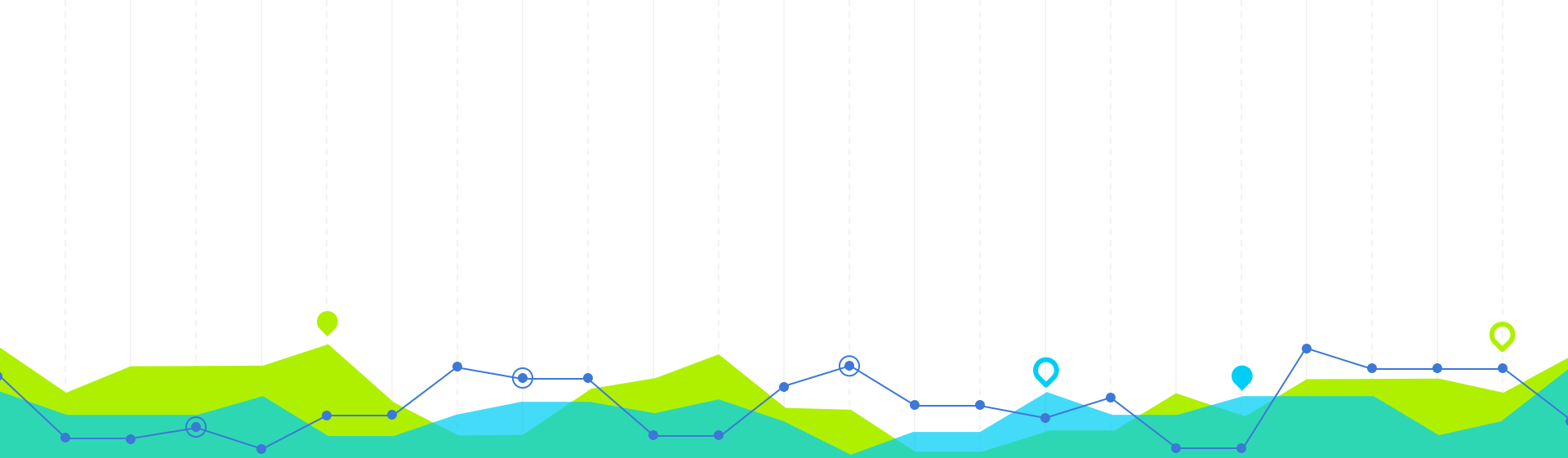


	Accuracy	Precision	Recall	F1 Score
Random Forest	0.9034612407	0.6879310345	0.3657195234	0.4775583483
Gradient Boosting	0.895	0.62	0.341	0.44
Logistic Regression	0.8866526595	0.5825	0.2135655362	0.3125419182
Naive Bayes	0.8897080507	0.543220339	0.401126408	0.4614830814

Results

- From our analysis, the best technique would be Random Forest
- It has the best accuracy, precision, and F1 score with Recall going to Naive Bayes
- That means the Random Forest has the best overall correctness, precision of positive predictions, and the balance between precision and recall although it may struggle to identify all positive instances correctly.





Thank you
For Listening