Acquisition Analytics – Marketing Campaign

Findings and the Methodology

By: Rashmi Singh

Objectives

The business objective of the bank is to capture 80% of total responders at the minimum possible cost. By building a response model based on the bank's previous marketing campaign dataset; we will predict the probability of a response from each prospect and target the ones most likely to respond for our next telemarketing campaign.

The steps were as follows:

- Identifying relevant predictor variables for a response using EDA.
- Build predictive models and choose the best one.
- Sorting the prospects in order of decreasing probability of response (predicted by the best model) and targeting the top 80%.

Analysis Approach

- The dataset contain round 41188 data points.

- Data cleaning, data preparation & EDA were performed to visualize the numerical and categorical features. We created dummy variables for the categorical features.

- RFE method has been used to select the variables with initial 18 variables, further repetitive model building has been done by removing features one at a time for high p-value and high VIF value . For each model confusion matrix was created to check precision, sensitivity, specificity and accuracy of the model.

- Precision-recall and ROC curve for model evaluation was done.

- The most important variables were identified for the given data set.

- To calculate cost of call we assumed $1 per Minute. (Cost = Duration/60)

- Created deciles to capture top 80% prospects & check the gain and Cumulative lift.

Methodology
## The ROC curve & Precision Recall curve
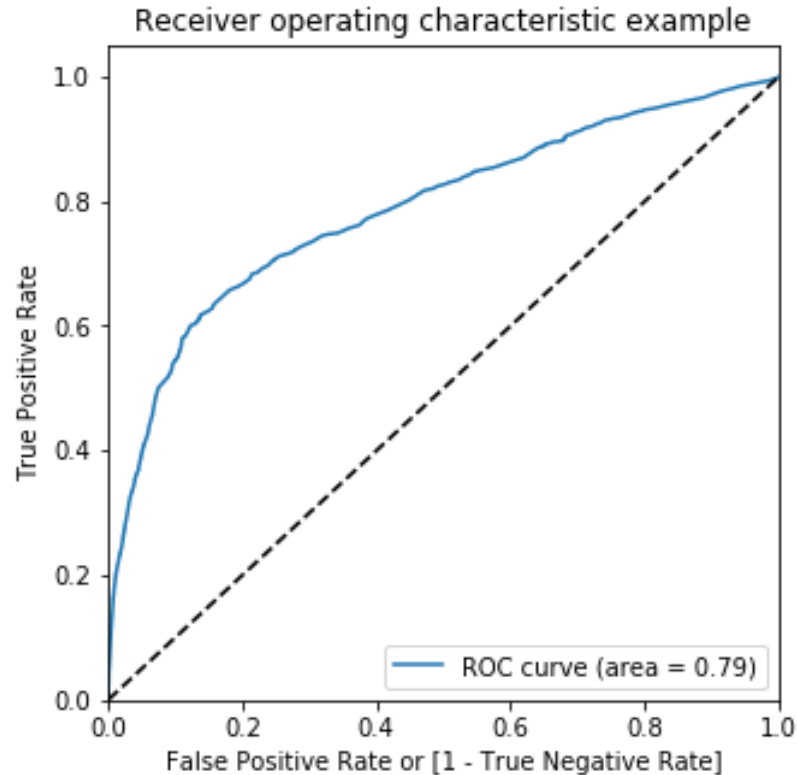
As the area under the ROC curve is higher i.e. 0.79, stating that our model is good.
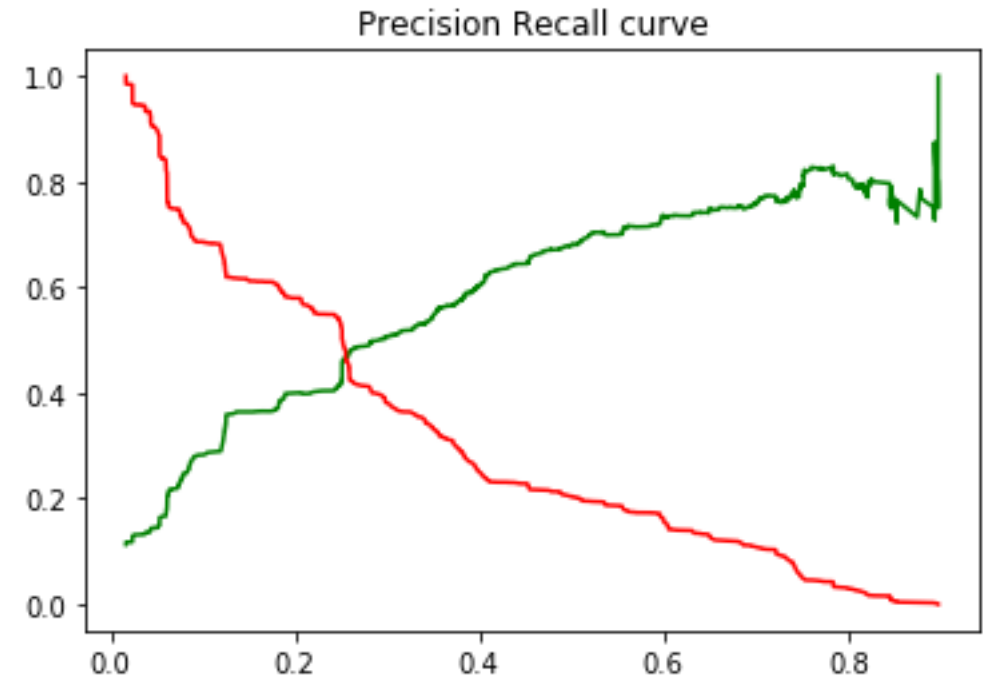
Based on the precision recall curve, we have taken a cutoff of 0.28

Deciles – Gain & Cumulative lift

| decile	total	Actual_Response	duration	Call_Cost	cumresp	gain	cumlift	cum_duration	cumcost	avg_duration
1	4116	1961	1182580	19709.36	1961	42.0	4.20	1182580	19709.36	287.31
2	3957	892	1073718	17895.07	2853	61.0	3.05	2256298	37604.43	279.49
3	4274	399	1099886	18331.33	3252	70.0	2.33	3356184	55935.76	271.82
4	3882	206	931404	15523.45	3458	75.0	1.88	4287588	71459.21	264.19
5	3902	243	1075661	17927.27	3701	80.0	1.60	5363249	89386.48	266.42
6	4074	216	952906	15881.80	3917	84.0	1.40	6316155	105268.28	260.94
7	4600	233	1055716	17595.32	4150	89.0	1.27	7371871	122863.60	255.92
8	4131	233	965891	16098.65	4383	94.0	1.18	8337762	138962.25	253.15
9	4034	110	1016312	16938.66	4493	97.0	1.08	9354074	155900.91	253.02
10	4218	147	1122062	18700.90	4640	100.0	1.00	10476136	174601.81	254.35
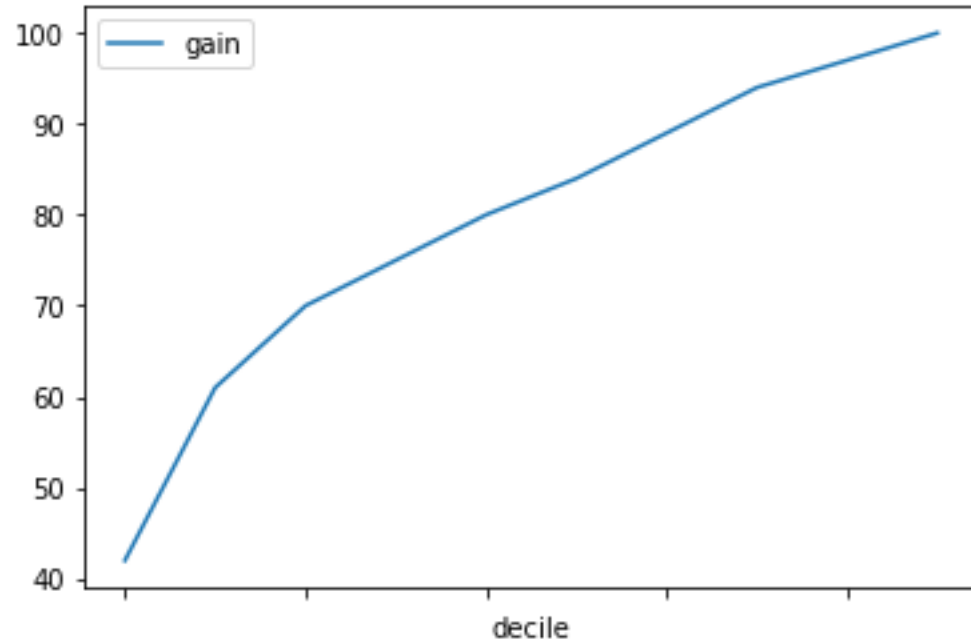
Gain & lift Chart

Gain chart tells us the number of responders captured (y-axis) as a function of the number of prospects targeted (x-axis).

In the 5th decile (50% people targeted), we can capture about 80% of the responders.
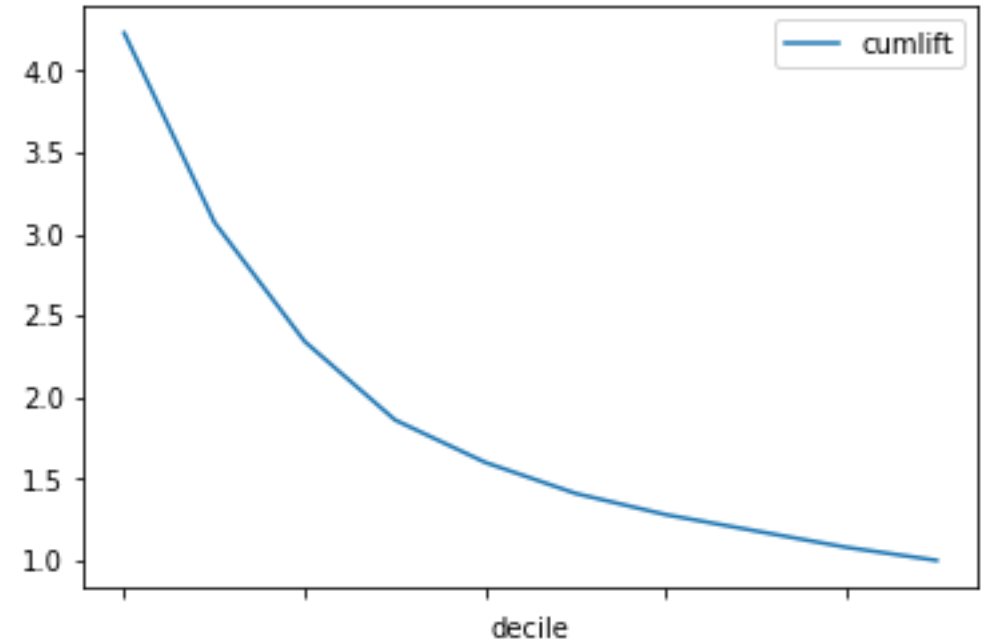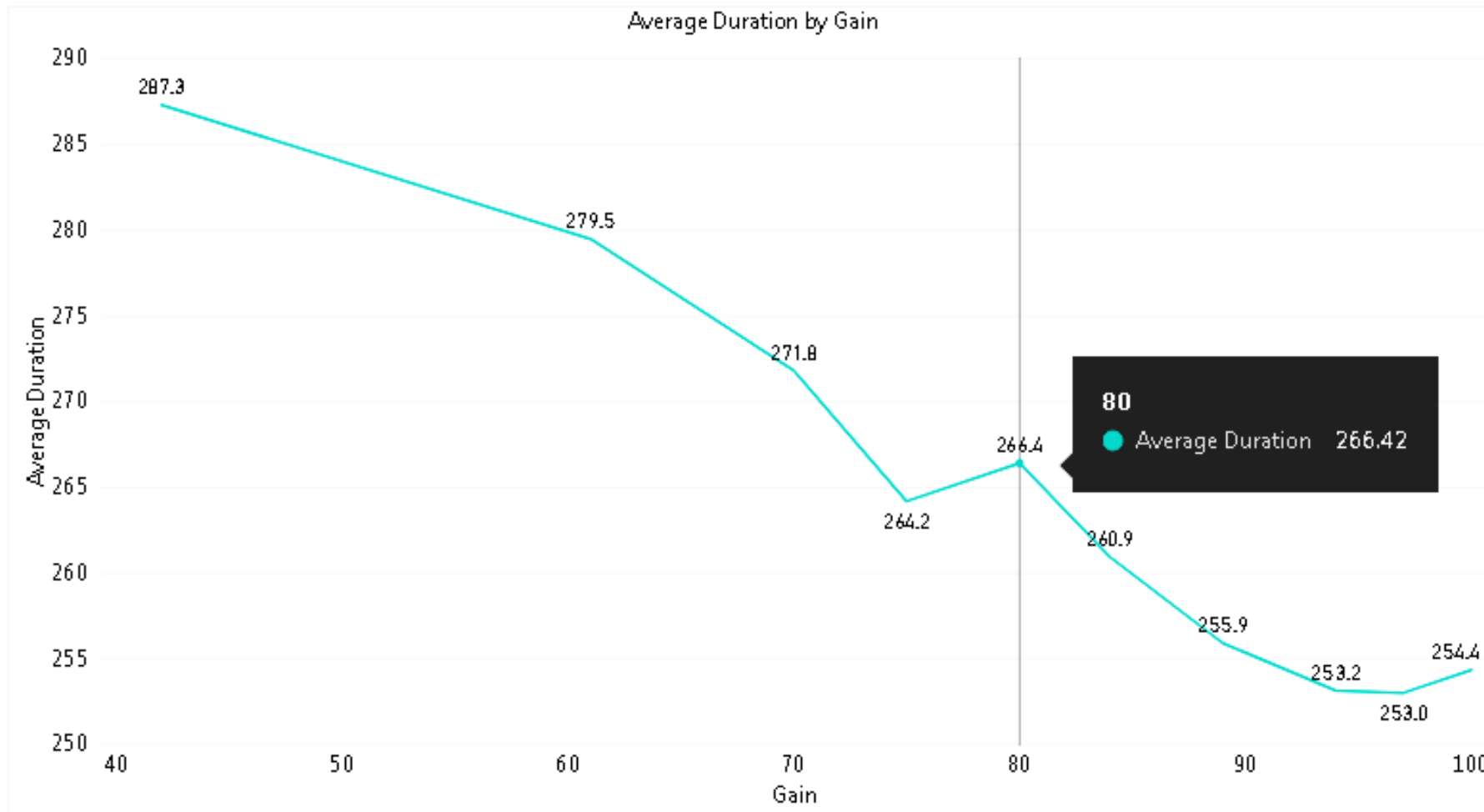
Lift chart compares the 'lift in response rate' we will get using the model viz-à-viz when we target the entire population (without using the model)

Findings - Top Deciles
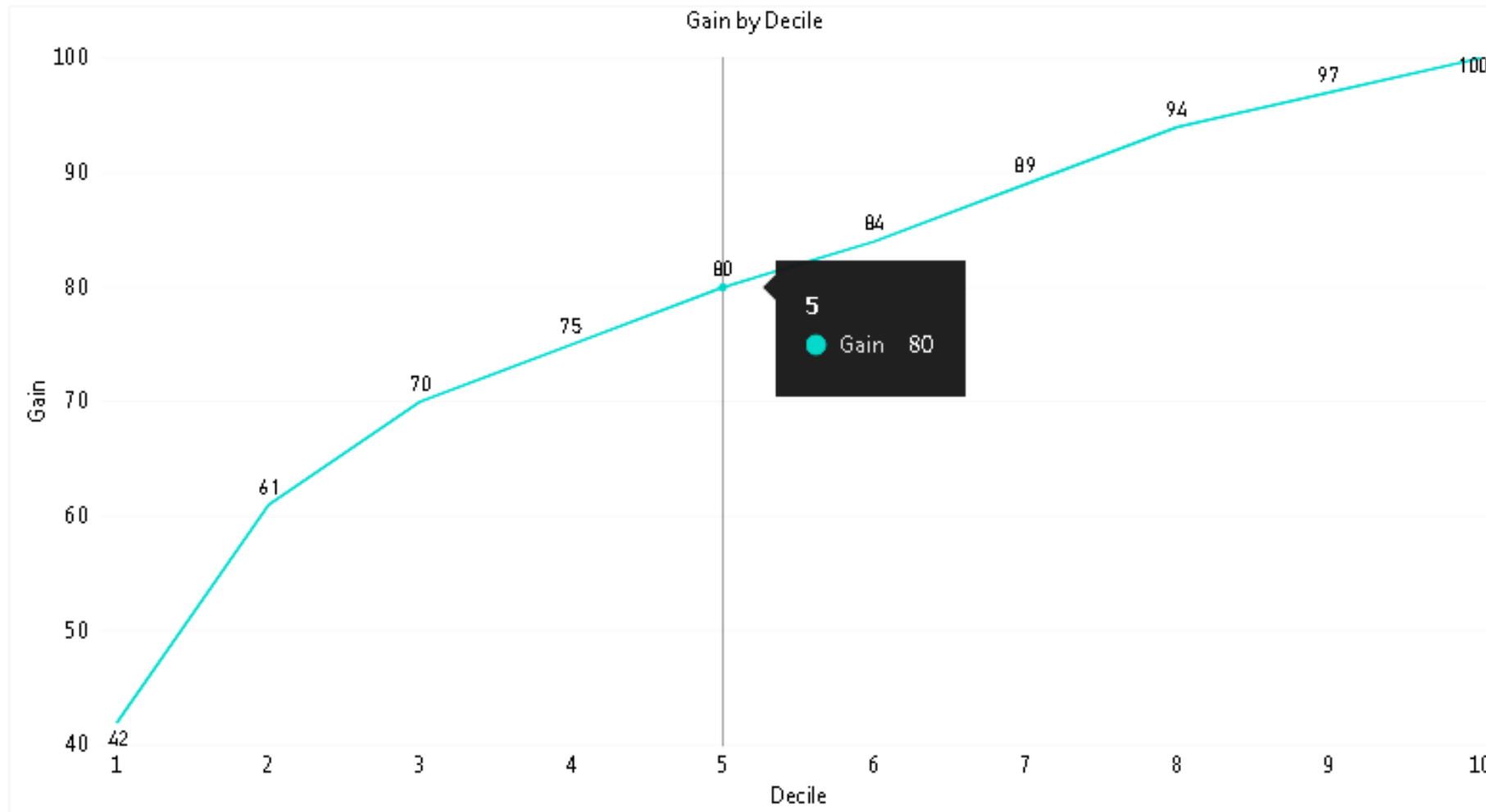
If we market to only the top **5 deciles** (50% of the customers), we will capture around 80% of the responders.
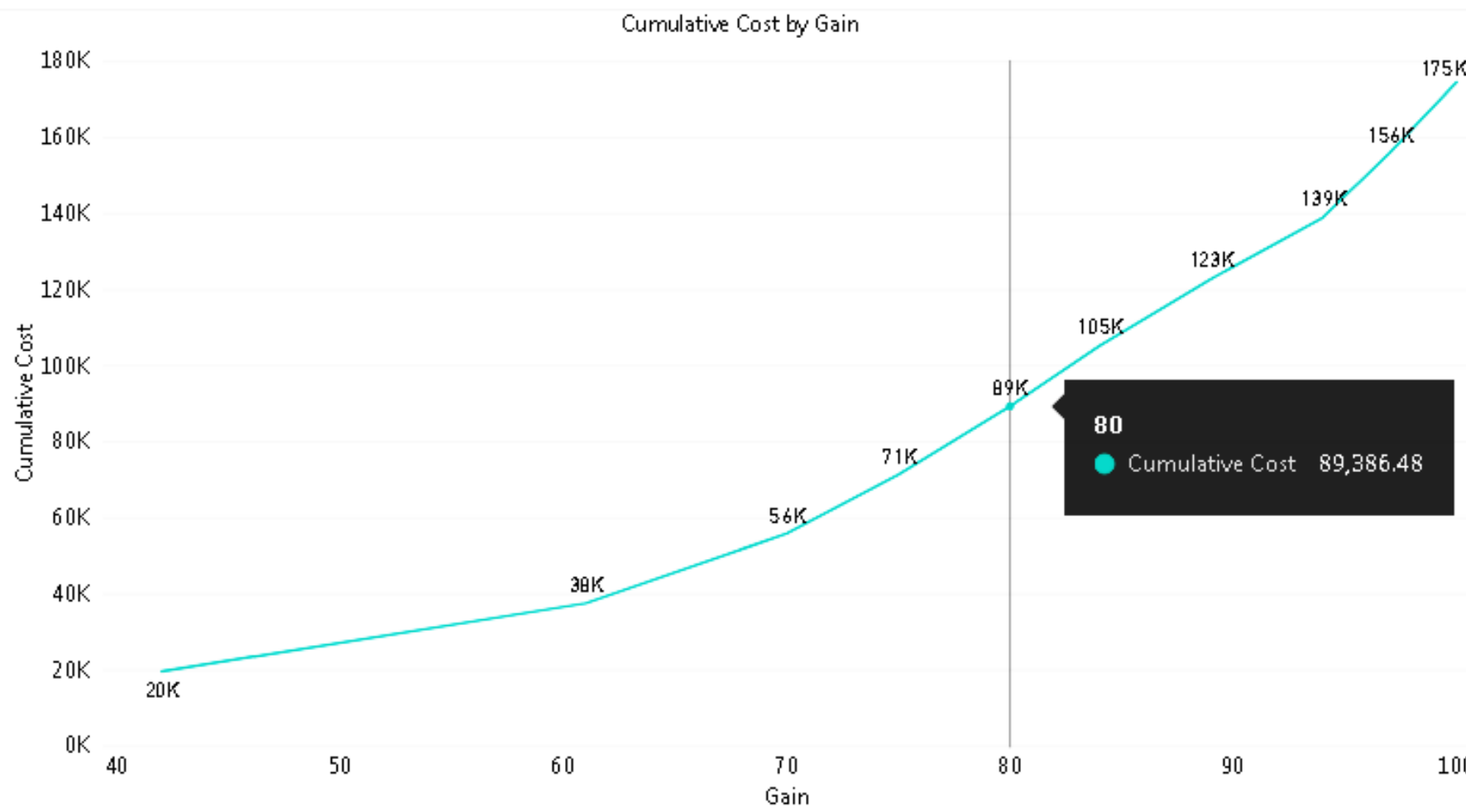
Findings - Average Call Duration

The average call duration for targeting the top 80% prospects is **266.42** seconds i.e. approx 4.5 minutes

Findings - Cost of acquisition

Cumulative Cost by Gain

The cost of acquiring 80% of the prospects using the predictive model will be **$89386.48.**

Thank You!