Q1

[54]:
```
print(f'beta_0 = {round(estimation_intercept[0],4)}')
for i in range(1,9):
    print(f'beta_{i} = {round(estimation_betas[0][i-1],4)}')
```

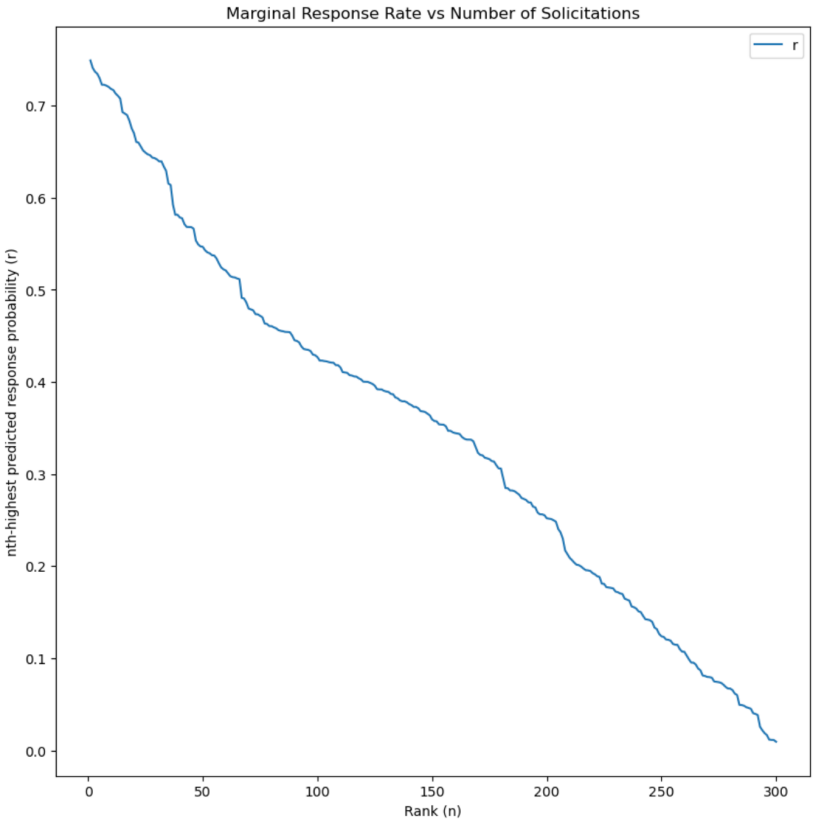
```
beta_0 = -10.2924
beta_1 = -0.0424
beta_2 = 0.0059
beta_3 = 0.0434
beta_4 = 0.0695
beta_5 = 0.0337
beta_6 = 0.0089
beta_7 = -0.001
beta_8 = -0.0009
```

The final equation is:

$t = -10.2924 - 0.0424Gender + 0.059FrequencyFoodWebsites + 0.0434FrequencyTravelWebsites + 0.0695RestaurantExp + 0.0337TravelExp + 0.0089$

t= -10.2924 - 0.0424 Gender + 0.059 FrequencyFoodWebsites + 0.0434 FrequencyTravelWebsites + 0.0695 RestaurantExp + 0.0337 TravelExp + 0.0089 EntertainmentExp - 0.0001 Income - 0.0009Ethnicity

Q4



Q5

Q5. Suppose average LT Customer equity is USD 30, and the solicitation cost is USD 12. Use the marginal cost rule to calculate how many of the best prospects in the holdout list Melrose should contact. Hand-in: your calculations and final answer.

The value of "p" is determined by dividing the solicitation cost by the equity. In this scenario, the solicitation cost is USD 12, and the equity is USD 30. Therefore, "p" must be greater than the quotient of 12 divided by 30, which is greater than 0.4.

$p_n > \frac{solicitation\_cost}{equity}$

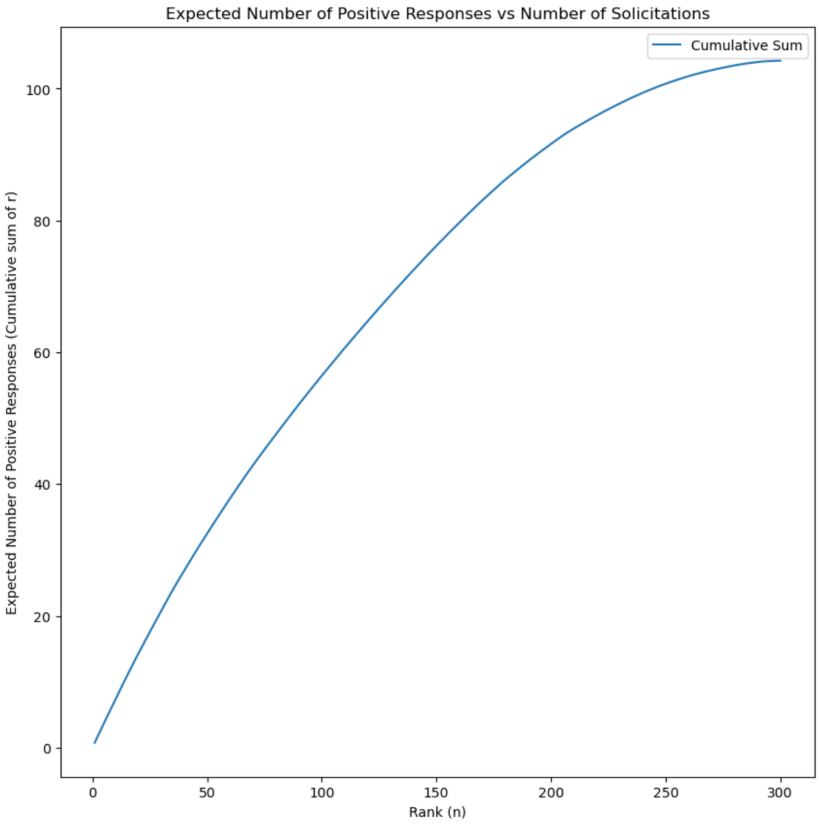
$p_n > \frac{12}{30}$

$p_n > 0.4$

[41]:
```
num_prospects = len(scoring_list[scoring_list["r"]>0.4])
print("They should contact", num_prospects, "prospects from the holdout list")
```

```
They should contact 122 prospects from the holdout list
```

Q6



The curve depicting the expected number of positive responses versus the number of solicitations does not rise as quickly as might be expected under the typical 80/20 rule due to the distribution and quality of the prospect data. If the top-ranked prospects (those with the highest response probabilities) do not have significantly higher probabilities than the average, the cumulative gains in expected responses will be gradual. This can happen if the predictive model is not highly discriminative or if the underlying prospect data does not vary significantly in terms of features that influence the response probability.

Q7

Q7. Assume that the firm has enough raw Belgian chocolate to make only 40 boxes of the pralines offered as part of the introductory offer. Use the limited supply rule to calculate how many of the best prospects in the holdout list Melrose should send solicitations to.

[46]:
```
best_prospects = scoring_list[scoring_list["Cumulative Sum"]<40]
best_prospects
```

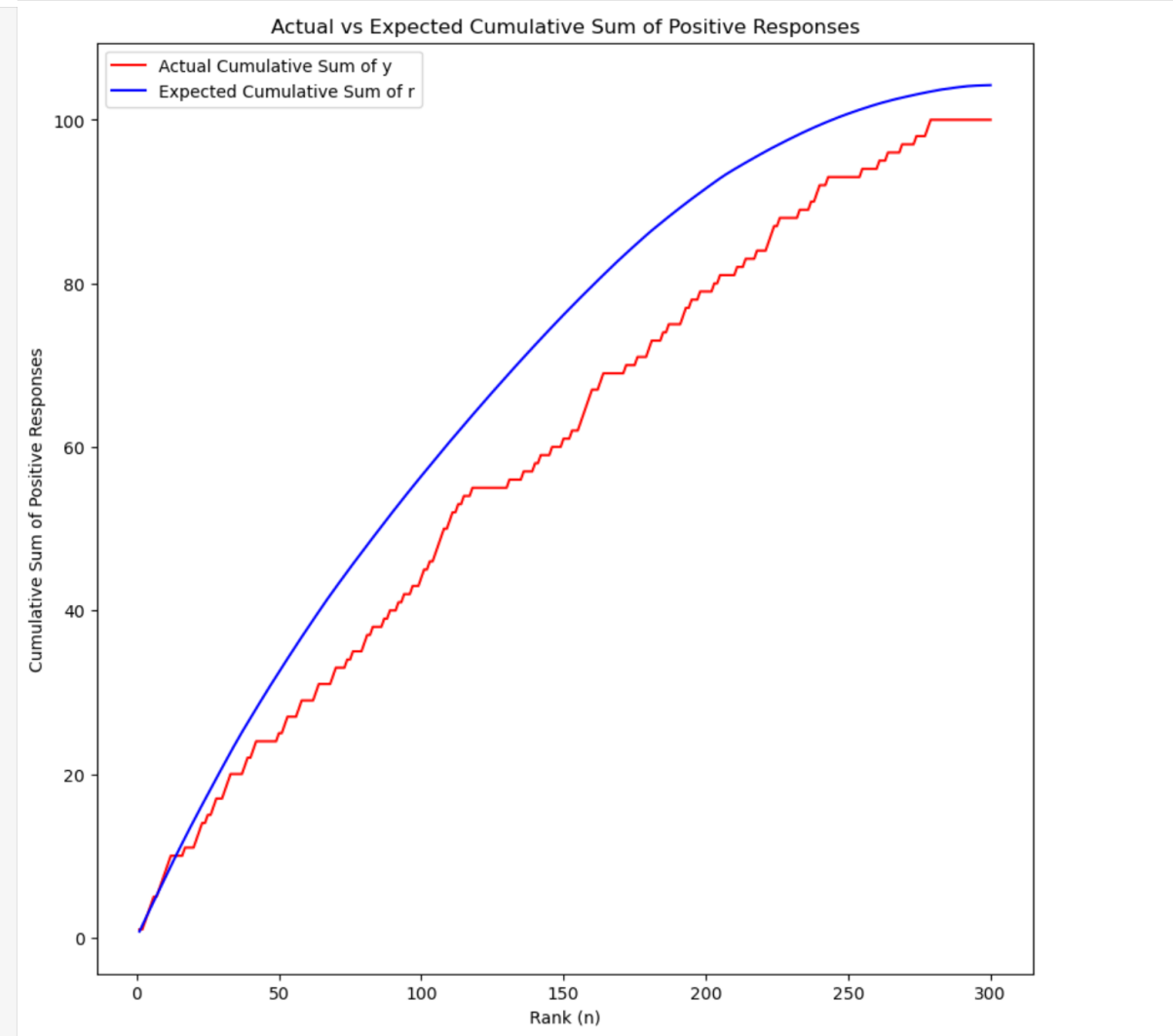
[46]:

	Gender	foodwebsites	travelwebsites	restaurantexp	travelexp	entertainmentexp	incm	ethnicdivneigh	Score	r	Lift	Rank	Cumulative Sum
131	1	317	57	70	63	12	26	0	1.094448	0.749218	2.081162	1	0.749218
243	0	321	57	70	63	0	28	0	1.051670	0.741095	2.058599	2	1.490314
191	0	315	57	70	63	0	0	14	1.030543	0.737021	2.047281	3	2.227335
159	0	311	57	70	63	0	0	0	1.018904	0.734759	2.040997	4	2.962094
19	1	314	57	70	63	0	0	0	0.994297	0.729936	2.027600	5	3.692030
...	...	...	...	...	...	...	...	...	...	...	...	...	...
91	1	163	57	70	63	0	0	15	0.084783	0.521183	1.447731	60	37.803171
257	0	129	57	70	63	15	0	0	0.072113	0.518021	1.438946	61	38.321191
55	1	159	57	70	63	0	15	0	0.059755	0.514934	1.430373	62	38.836125
143	0	254	57	60	63	12	13	28	0.055611	0.513899	1.427498	63	39.350024
178	0	211	57	70	52	0	0	0	0.054031	0.513504	1.426401	64	39.863529

64 rows × 13 columns

[47]:
```
print("they should contact", len(best_prospects), "prospects in the holdout list given only 40 boxes in hand")
```
```
they should contact 64 prospects in the holdout list given only 40 boxes in hand
```

Q8



Discrepancies between the actual responses and the predicted probabilities could indicate several issues:
1. Model Performance: The predictive model might not be capturing all the nuances of the data, leading to over- or under-estimation of certain prospects' likelihood to respond.
2. Data Quality: If there are issues with the data quality or relevant features are not included in the model, the predictions may not align well with actual outcomes.
3. External Factors: Factors not accounted for in the model, such as external economic conditions or competitors' promotions, could affect the actual response rates.

The impact of these discrepancies on decision-making is significant, especially when resources are limited. Overestimating the likelihood of response could lead to wasteful expenditure on prospects less likely to convert, while underestimating could mean missing out on potential conversions. In scenarios where only a limited number of products (like the 40 boxes of chocolates) are available, it's crucial that the predictive model is accurate to ensure that solicitations are sent to those most likely to respond, maximizing revenue and minimizing costs.

Q9

The Pareto Principle, often referred to as the 80-20 rule, suggests that 80% of outcomes (or outputs) result from 20% of all causes (or inputs) for any given event. In a business context, a common interpretation of the Pareto Principle is that 80% of sales come from 20% of clients.

Statement:
To prove that if the Pareto 80-20 phenomenon holds, then the average lift in the top 20% of the prospects will be 4.

Given:
P prospects in total.
C total conversions.
- 80% of conversions come from 20% of the prospects.

Definitions:
- Let P_{20} be the top 20% of prospects. So, $P_{20} = 0.2P$
- Let C_{80} be the conversions from the top 20% of prospects. So, $C_{80} = 0.8C$

Average Conversion:
- The average conversion rate across all prospects is C/P
Conversion Rate for Top 20%:
- The conversion rate for the top 20% of prospects is $C_{80}/P_{20} = 4C/P$
Proof:
The lift is a ratio of the conversion rate of a group to the average conversion rate. Thus, the average lift in the top 20% of prospects, compared to the overall average, is calculated as:

$\text{Lift} = (4C/P)/(C/P) = 4$

This concludes the proof that under the Pareto 80-20 rule, the average lift in the top 20% of prospects is 4, meaning the conversion rate among these prospects is 4 times higher than the overall average conversion rate.