ANALYSIS OF TELECOM

MODELING CUSTOMER DATA TO REDUCE CUSTOMER CHURN

OVERVIEW

▸ Current customer data analysis

▸ Cleaning and structuring of data to make it accessible for analytics

▸ Creation of models to accurately predict future customer churn

ANALYSIS OF CURRENT CUSTOMER DATA

▸ Key metrics identified: Total Minutes of phone plans used, customer service calls

▸ Very specific cases cause customer churn
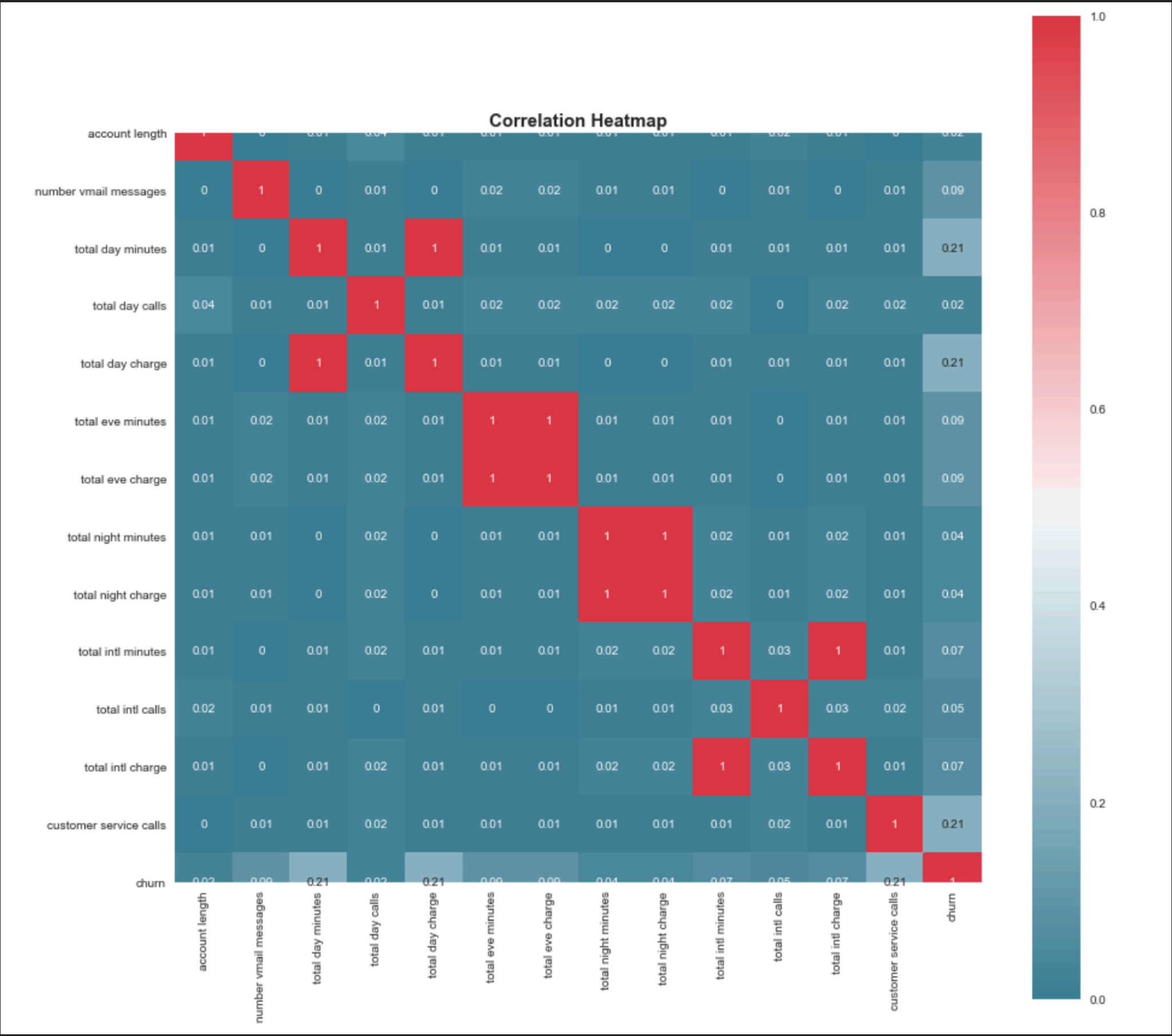
▸ Overall low correlation numbers with customer churn

Charts showing where churn is most prevalent
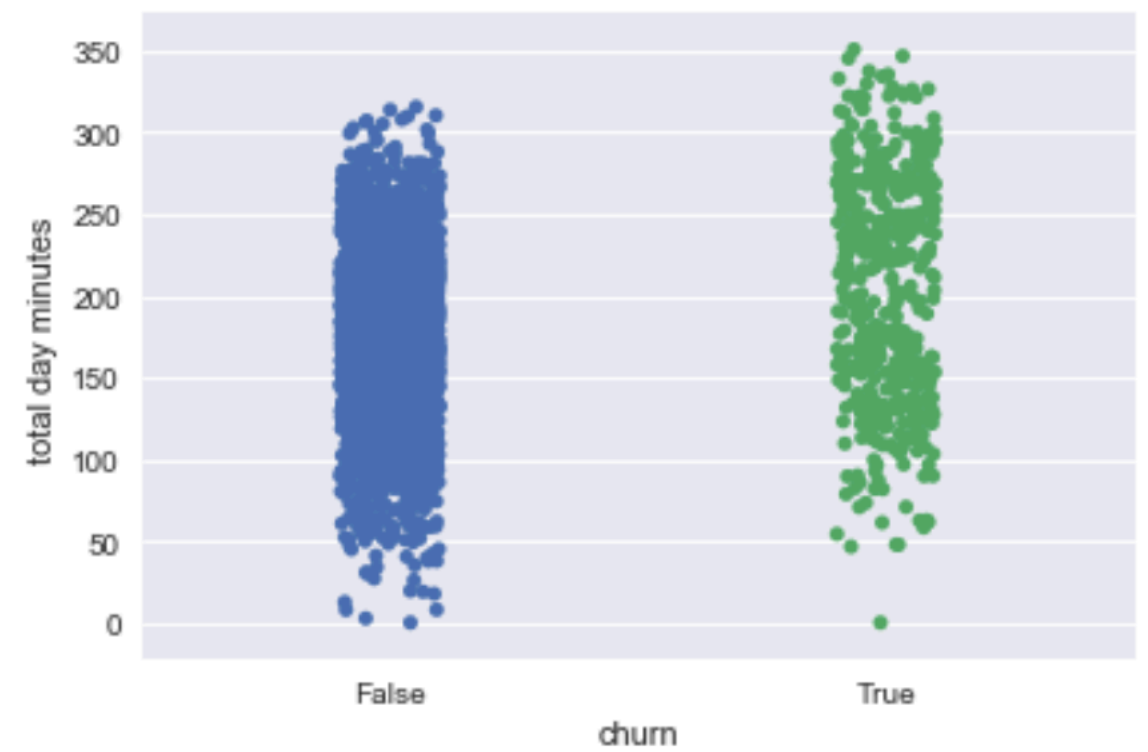
SORTING AND CLEANING DATA

▸ Data containing customer information removed

▸ Data was analyzed and sorted to look for and remove null/ missing values

▸ Highest correlation columns were included in the final data

Correlation Heatmap

Correlation Heatmap

BUILDING MODELS TO PREDICT CHURN ON FUTURE DATA

▸ 3 Models were chosen for this project: K Nearest Neighbor, Grid Search, XG Boost.

▸ Even without model optimization each model produced a score above 90% accuracy.

▸ XGboost returned the highest accuracy improvements after multiple iterations

▸ KNN returned the best initial model results post data scaling.

As seen in this chart, returned from the KNN model customer churn numbers are represented in the last column when seen as a ~2.44 value indicate customer churn. This model was scored as having ~98% accuracy.

	account length	number vmail messages	total day minutes	total day calls	total day charge	total eve minutes	total eve charge	total night minutes	total night charge	total intl minutes	total intl calls	total intl charge	customer service calls	churn
0	-1.404508	-0.584700	-1.883677	1.330852	-1.884170	1.037727	1.037905	1.069609	1.071757	2.768109	-0.595235	2.767357	-0.427903	-0.408915
1	0.366388	-0.584700	0.294083	0.529165	0.293703	0.516178	0.517286	2.214376	2.212569	1.051970	-1.417963	1.045965	0.324616	-0.408915
2	0.518179	1.685101	1.056392	-1.875896	1.056666	0.093407	0.094283	-0.077125	-0.077797	-0.056369	0.227493	-0.053077	1.829653	-0.408915
3	2.010792	-0.584700	-0.679156	1.681590	-0.679320	-0.402459	-0.403094	-0.322994	-0.322569	-0.056369	-0.183871	-0.053077	-0.427903	-0.408915
4	0.290493	-0.584700	0.484660	1.080325	0.484172	-0.718549	-0.719184	-1.186487	-1.188013	0.837453	1.050221	0.834102	1.077134	-0.408915
5	-0.696150	-0.584700	1.141504	0.729587	1.141561	-0.639527	-0.640162	-0.606236	-0.606679	0.301160	-1.006599	0.304443	0.324616	-0.408915
6	1.985493	2.197637	0.741847	-1.124315	0.742121	0.660394	0.661386	0.721458	0.722083	-0.056369	-0.183871	-0.053077	1.829653	-0.408915
7	0.164000	2.197637	0.447655	-2.176529	0.447166	0.099334	0.098931	1.689199	1.688058	-0.699921	-0.183871	-0.701909	-1.180421	-0.408915
8	-0.594956	-0.584700	0.175666	0.579270	0.176157	0.121065	0.122173	0.265125	0.263135	-0.378145	1.050221	-0.384114	-1.180421	-0.408915
9	0.062806	-0.584700	-0.248044	-0.823682	-0.248316	-0.167367	-0.168351	-0.643608	-0.641647	-0.914438	0.638857	-0.913773	-0.427903	-0.408915
10	0.619373	-0.584700	-1.367454	-0.272522	-1.367184	-0.854863	-0.853987	-0.159737	-0.160845	-0.771427	-0.595235	-0.768117	-0.427903	-0.408915
11	-0.569657	-0.584700	0.534617	0.930008	0.534238	1.017971	1.016987	1.565281	1.565672	0.908959	1.050221	0.913550	-1.180421	-0.408915
12	-0.291373	0.293932	0.147912	0.228532	0.147858	-0.979324	-0.979494	0.025156	0.027105	0.086642	-0.595235	0.092579	0.324616	-0.408915
13	-0.291373	1.977978	-0.103724	0.378848	-0.103560	-0.082418	-0.082356	-1.440224	-1.441526	0.551430	-0.595235	0.556031	-0.427903	-0.408915
14	-0.645553	-0.584700	0.579024	-1.976107	0.578862	-1.882156	-1.881280	-0.486252	-0.484293	-0.199381	-0.183871	-0.198733	1.829653	2.445495
15	0.366388	-0.584700	-0.967797	-1.675475	-0.967744	-0.355046	-0.354286	-0.545260	-0.545486	0.122395	1.050221	0.119062	-0.427903	-0.408915
16	-0.038388	-0.584700	1.130402	-1.424947	1.130677	-0.894375	-0.893499	-0.244316	-0.243892	0.479924	2.284313	0.476582	-0.427903	-0.408915
17	0.897657	0.586810	-1.657945	0.228532	-1.657784	0.008458	0.008288	0.520829	0.521020	1.016217	-0.183871	1.019482	-1.180421	-0.408915
18	-0.847941	1.977978	1.209964	-0.573155	1.210129	-1.040566	-1.039923	1.443330	1.443286	-0.985944	0.638857	-0.979980	-0.427903	-0.408915
19	0.998851	-0.584700	1.026788	-0.973998	1.027279	1.112798	1.112279	-1.918194	-1.917957	0.444171	-0.183871	0.450099	0.324616	-0.408915

CONCLUSIONS

▸ Based on the models developed in the project it is clear that modeling these churn data-sets can be highly effective at predicting which customers are at risk of churn.

▸ Telecom should look at improving and reducing customer service tickets as those were one of the primary indicators of customer churn.

▸ Additionally creating a flat-rate or packaged billing system for high volume uses could also reduce customer churn.