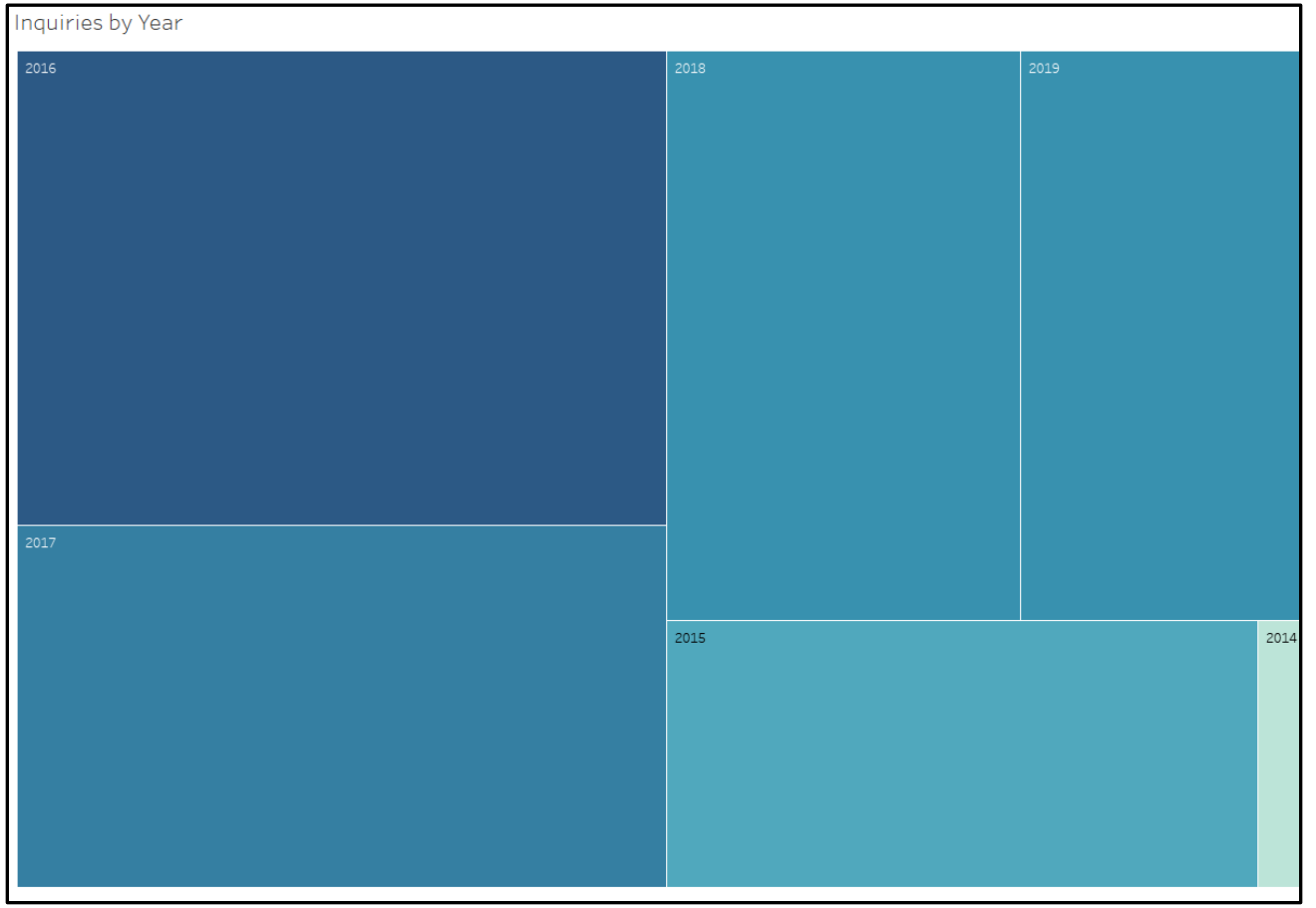
Problem Sets

We decided to analyze project types and which types tend to show higher performance results.

- Which project types produced the best outcomes in survey data?
- Which project types perform better on the platform by match rate, days to match, and number of inquiries received?
- Which project types take the longest to complete?



Datasets

- We used four data sets from taproot to get the desired results: Project_export, Project_categories, Project_Inquiries and Session_export.
- We cleaned the data using python to remove NULL values from the columns and rows to reduce redundancy and dimensionality.
- We created new calculated columns such as days_to_update (seen below) and succeeded by mapping existing values and grouping them to category keys.

```
In [14]: Prjts.sort_values(by='id',axis=0).head()
```

Out[14]:

id	organization_id	description	created_at	updated_at	state	user_id	project_inquiries_count	project_category_id	timeline	days_to_update
31	35	multipage brochure design	2014-08-19 14:53:23.622273	2015-02-20 19:59:24.590494	completed	1301	1	14	2	185
32	36	print design	2014-08-19 14:53:23.661564	2015-02-20 19:59:24.614798	completed	1302	1	14	2	185
33	37	NaN	2014-08-19 14:53:23.668844	2015-06-28 02:48:38.165214	completed	1303	1	20	2	312
34	477	Marketing consultation for campaign design	2014-08-19 14:53:23.683657	2015-02-20 19:59:24.628209	completed	1304	1	10	2	185
35	39	photography	2014-08-19 14:53:23.697066	2015-02-20 19:59:24.639884	closed	1305	0	13	2	185

```
In [21]: User_Inqry = User_Inqry.drop(labels=['scheduled_for', 'decision_deadline', 'conference_line_id', 'hours', 'pbc_rating',
                                              'npo_rating', 'satisfaction_rating', 'pbc_review', 'archived', 'time_slots'], axis=1)
         User_Inqry['created_at'] = pd.to_datetime(User_Inqry['created_at'], format="%Y-%m-%d %H:%M:%S.%f")
         User_Inqry['updated_at'] = pd.to_datetime(User_Inqry['updated_at'], format="%Y-%m-%d %H:%M:%S.%f")
         User_Inqry['days_to_update'] = (User_Inqry['updated_at']-User_Inqry['created_at']).dt.days

In [22]: User_Inqry['state'].unique()

Out[22]: array(['rejected', 'pbc_expired', 'npo_expired', 'cancelled', 'completed',
                'failed', 'accepted', 'admin_close', 'npo_rescheduled',
                'confirmed', 'applied', 'pbc_rescheduled', 'missed'], dtype=object)

In [23]: mapping={
             'rejected': 'inqry_failed',
             'pbc_expired': 'inqry_failed',
             'npo_expired': 'inqry_failed',
             'cancelled': 'inqry_failed',
             'failed': 'inqry_failed',
             'admin_close': 'inqry_failed',
             'missed': 'inqry_failed'
         }
         User_Inqry['inqry_failed'] = User_Inqry['state'].map(mapping) == 'inqry_failed'
         User_Inqry["inqry_failed"] = User_Inqry["inqry_failed"].astype(int)
```

```
In [24]: User_Inqry.tail()
```

Out[24]:

	id	user_id	project_id	qualifications	created_at	updated_at	state	days_to_update	inqry_failed
25079	26952	178353	11395	I have design, content generation, and proofre...	2019-11-20 22:00:51.252953	2019-12-11 19:59:09.406476	pbc_expired	20	1
25080	26920	4172	11396	I live in Silicon Valley and work in tech as w...	2019-11-19 12:00:24.619799	2019-11-28 00:35:09.857962	accepted	8	0
25081	27064	182933	11450	I began my post-graduate professional career w...	2019-11-28 04:14:16.024452	2019-11-28 22:11:00.901167	rejected	0	1
25082	27033	174579	4129	I am an integrated communications expert with ...	2019-11-25 23:09:52.848121	2019-11-28 23:42:01.560910	accepted	3	0
25083	27069	182947	11796	Hello,\nMy background is in Business Analysis ...	2019-11-28 17:41:20.908411	2019-12-10 03:57:53.690976	accepted	11	0

- Here we can see an example of some of the data preprocessing of the user_inqry dataset and a glance of the data set tail as a sanity check that the data looks valid.

Tools Used and Visualization

- We are using Python and Jupyter Notebooks to clean and process the data due to its flexibility and ease of documentation.
- Tableau is used for data visualization due to its user friendly and intuitive GUI and its beautiful visuals.

Q.1 Which project types produce the best outcomes in survey data?

- We can see from the graph that Copy writing/editing has on average the best outcomes with a 59% success rate among all the project categories.
- Interestingly, most project categories had an average success rate below 50%, with the marketing group category taking both the top and bottom spots with a 29% success rate of multimedia development alongside copy writing/editing in the top spot.

Q.2 Which project types perform better on the platform? Match rate, days to match, # of inquiries received etc.

- Of note, it doesn't appear that there is a correlation between number of volunteer inquiries received and the days taken to match a volunteer to a project.
- Most volunteer inquiries do not result in being matched with a project. Regardless of category, volunteers have between a 65-80% chance of not working on a project.

Q.3 Which project types take the longest to complete? Shortest?

- We can see from the scatter plot that Public Relations takes longest time to complete with an average of 178 days.
- Staff Development takes the shortest time with 99 days on average.
- Across all categories a project takes on average 140 days.
- Or, put another way, each project takes on average just over half a year to complete from the moment a project is posted.