MACHINE LEARNING : Classification Model for Citizen engagement in Smart-City.

Mid-Bootcamp Project

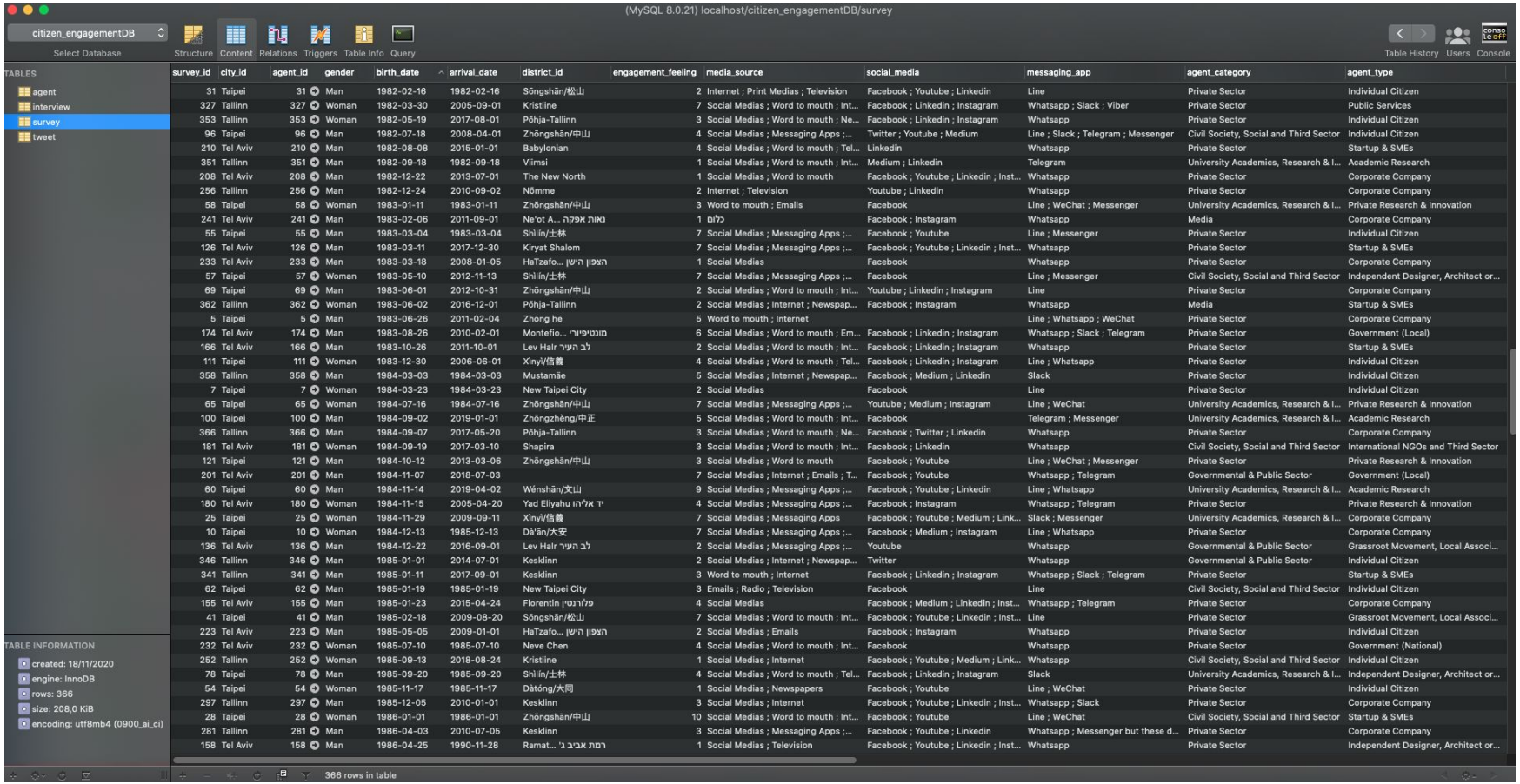
IRONHACK BERLIN 20th Nov. 2020

General :

A statistical classification model identifying categories of citizen engagement and ranking cities on their citizen engagement dynamic.

So far, a citizen highly engaged in the digital transformation of its city will be defined as :
- using social medias a source to form its opinion
- meeting peers online
- sharing its opinion easily in public
- willing to engagement more
- notably by online survey studies and voting apps

Creating database : citizen_engagementDB in SQL



survey_id	city_id	agent_id	gender	birth_date	arrival_date	district_id	engagement_feeling	media_source	social_media	messaging_app	agent_category	agent_type
31	Taipei	31	Man	1982-02-16	1982-02-16	Sōngshān/松山	2	Internet ; Print Medias ; Television	Facebook ; Youtube ; Linkedin	Line	Private Sector	Individual Citizen
327	Tallinn	327	Woman	1982-03-30	2005-09-01	Kristiine	7	Social Medias ; Word to mouth ; Int...	Facebook ; Linkedin ; Instagram	Whatsapp ; Slack ; Viber	Private Sector	Public Services
353	Tallinn	353	Woman	1982-05-19	2017-08-01	Põhja-Tallinn	3	Social Medias ; Word to mouth ; Ne...	Facebook ; Linkedin ; Instagram	Whatsapp	Private Sector	Individual Citizen
96	Taipei	96	Man	1982-07-18	2008-04-01	Zhōngshān/中山	4	Social Medias ; Messaging Apps ;...	Twitter ; Youtube ; Medium	Line ; Slack ; Telegram ; Messenger	Civil Society, Social and Third Sector	Individual Citizen
210	Tel Aviv	210	Man	1982-08-08	2015-01-01	Babylonian	4	Social Medias ; Word to mouth ; Tel...	Linkedin	Whatsapp	Private Sector	Startup & SMEs
351	Tallinn	351	Man	1982-09-18	1982-09-18	Viimsi	1	Social Medias ; Word to mouth ; Int...	Medium ; Linkedin	Telegram	University Academics, Research & I...	Academic Research
208	Tel Aviv	208	Man	1982-12-22	2013-07-01	The New North	1	Social Medias ; Word to mouth	Facebook ; Youtube ; Linkedin ; Inst...	Whatsapp	Private Sector	Corporate Company
256	Tallinn	256	Man	1982-12-24	2010-09-02	Nõmme	2	Internet ; Television	Youtube ; Linkedin	Whatsapp	Private Sector	Corporate Company
58	Taipei	58	Woman	1983-01-11	1983-01-11	Zhōngshān/中山	3	Word to mouth ; Emails	Facebook	Line ; WeChat ; Messenger	University Academics, Research & I...	Private Research & Innovation
241	Tel Aviv	241	Man	1983-02-06	2011-09-01	Ne'ot A... נאות אפקה	1	כלום	Facebook ; Instagram	Whatsapp	Media	Corporate Company
55	Taipei	55	Man	1983-03-04	1983-03-04	Shilin/士林	7	Social Medias ; Messaging Apps ;...	Facebook ; Youtube	Line ; Messenger	Private Sector	Individual Citizen
126	Tel Aviv	126	Man	1983-03-11	2017-12-30	Kiryat Shalom	7	Social Medias ; Messaging Apps ;...	Facebook ; Youtube ; Linkedin ; Inst...	Whatsapp	Private Sector	Startup & SMEs
233	Tel Aviv	233	Man	1983-03-18	2008-01-05	HaTzafo... הצפון הישן	1	Social Medias	Facebook	Whatsapp	Private Sector	Corporate Company
57	Taipei	57	Woman	1983-05-10	2012-11-13	Shilin/士林	7	Social Medias ; Messaging Apps ;...	Facebook	Line ; Messenger	Civil Society, Social and Third Sector	Independent Designer, Architect or...
69	Taipei	69	Man	1983-06-01	2012-10-31	Zhōngshān/中山	2	Social Medias ; Word to mouth ; Int...	Youtube ; Linkedin ; Instagram	Line	Private Sector	Corporate Company
362	Tallinn	362	Woman	1983-06-02	2016-12-01	Põhja-Tallinn	2	Social Medias ; Internet ; Newspap...	Facebook ; Instagram	Whatsapp	Media	Startup & SMEs
5	Taipei	5	Man	1983-06-26	2011-02-04	Zhong he	5	Word to mouth ; Internet		Line ; Whatsapp ; WeChat	Private Sector	Corporate Company
174	Tel Aviv	174	Man	1983-08-26	2010-02-01	Montefio... מונטיפיורי	6	Social Medias ; Word to mouth ; Em...	Facebook ; Linkedin ; Instagram	Whatsapp ; Slack ; Telegram	Private Sector	Government (Local)
166	Tel Aviv	166	Man	1983-10-26	2011-10-01	Lev Halr לב העיר	2	Social Medias ; Word to mouth ; Int...	Facebook ; Linkedin ; Instagram	Whatsapp	Private Sector	Startup & SMEs
111	Taipei	111	Woman	1983-12-30	2006-06-01	Xinyi/信義	4	Social Medias ; Word to mouth ; Tel...	Facebook ; Linkedin ; Instagram	Line ; Whatsapp	Private Sector	Individual Citizen
358	Tallinn	358	Man	1984-03-03	1984-03-03	Mustamäe	5	Social Medias ; Internet ; Newspap...	Facebook ; Medium ; Linkedin	Slack	Private Sector	Individual Citizen
7	Taipei	7	Woman	1984-03-23	1984-03-23	New Taipei City	2	Social Medias	Facebook	Line	Private Sector	Individual Citizen
65	Taipei	65	Woman	1984-07-16	1984-07-16	Zhōngshān/中山	7	Social Medias ; Messaging Apps ;...	Youtube ; Medium ; Instagram	Line ; WeChat	University Academics, Research & I...	Private Research & Innovation
100	Taipei	100	Man	1984-09-02	2019-01-01	Zhōngzhèng/中正	5	Social Medias ; Word to mouth ; Int...	Facebook	Telegram ; Messenger	University Academics, Research & I...	Academic Research
366	Tallinn	366	Man	1984-09-07	2017-05-20	Põhja-Tallinn	3	Social Medias ; Word to mouth ; Ne...	Facebook ; Twitter ; Linkedin	Whatsapp	Private Sector	Corporate Company
181	Tel Aviv	181	Woman	1984-09-19	2017-03-10	Shapira	3	Social Medias ; Word to mouth ; Int...	Facebook ; Linkedin	Whatsapp	Civil Society, Social and Third Sector	International NGOs and Third Sector
121	Taipei	121	Man	1984-10-12	2013-03-06	Zhōngshān/中山	3	Social Medias ; Word to mouth	Facebook ; Youtube	Line ; WeChat ; Messenger	Private Sector	Private Research & Innovation
201	Tel Aviv	201	Man	1984-11-07	2018-07-03		7	Social Medias ; Internet ; Emails ; T...	Facebook ; Youtube	Whatsapp ; Telegram	Governmental & Public Sector	Government (Local)
60	Taipei	60	Man	1984-11-14	2019-04-02	Wénshān/文山	9	Social Medias ; Messaging Apps ;...	Facebook ; Youtube ; Linkedin	Line ; Whatsapp	University Academics, Research & I...	Academic Research
180	Tel Aviv	180	Woman	1984-11-15	2005-04-20	Yad Eliyahu יד אליהו	4	Social Medias ; Messaging Apps ;...	Facebook ; Instagram	Whatsapp ; Telegram	Private Sector	Private Research & Innovation
25	Taipei	25	Woman	1984-11-29	2009-09-11	Xinyi/信義	7	Social Medias ; Messaging Apps	Facebook ; Youtube ; Medium ; Link...	Slack ; Messenger	University Academics, Research & I...	Corporate Company
10	Taipei	10	Woman	1984-12-13	1985-12-13	Dà'ān/大安	7	Social Medias ; Messaging Apps ;...	Facebook ; Medium ; Instagram	Line ; Whatsapp	Private Sector	Corporate Company
136	Tel Aviv	136	Man	1984-12-22	2016-09-01	Lev Halr לב העיר	2	Social Medias ; Messaging Apps ;...	Youtube	Whatsapp	Governmental & Public Sector	Grassroot Movement, Local Associ...
346	Tallinn	346	Man	1985-01-01	2014-07-01	Kesklinn	2	Social Medias ; Internet ; Newspap...	Twitter	Whatsapp	Governmental & Public Sector	Individual Citizen
341	Tallinn	341	Man	1985-01-11	2017-09-01	Kesklinn	3	Word to mouth ; Internet	Facebook ; Linkedin ; Instagram	Whatsapp ; Slack ; Telegram	Private Sector	Startup & SMEs
62	Taipei	62	Man	1985-01-19	1985-01-19	New Taipei City	3	Emails ; Radio ; Television	Facebook	Line	Civil Society, Social and Third Sector	Individual Citizen
155	Tel Aviv	155	Man	1985-01-23	2015-04-24	Florentin פלורנטין	4	Social Medias	Facebook ; Medium ; Linkedin ; Inst...	Whatsapp ; Telegram	Private Sector	Corporate Company
41	Taipei	41	Man	1985-02-18	2009-08-20	Sōngshān/松山	7	Social Medias ; Word to mouth ; Int...	Facebook ; Youtube ; Linkedin ; Inst...	Line	Private Sector	Grassroot Movement, Local Associ...
223	Tel Aviv	223	Man	1985-05-05	2009-01-01	HaTzafo... הצפון הישן	2	Social Medias ; Emails	Facebook ; Instagram	Whatsapp	Private Sector	Individual Citizen
232	Tel Aviv	232	Woman	1985-07-10	1985-07-10	Neve Chen	4	Social Medias ; Word to mouth ; Int...	Facebook	Whatsapp	Private Sector	Government (National)
252	Tallinn	252	Woman	1985-09-13	2018-08-24	Kristiine	1	Social Medias ; Internet	Facebook ; Youtube ; Medium ; Link...	Whatsapp	Civil Society, Social and Third Sector	Individual Citizen
78	Taipei	78	Man	1985-09-20	1985-09-20	Shilin/士林	4	Social Medias ; Word to mouth ; Tel...	Facebook ; Linkedin ; Instagram	Slack	University Academics, Research & I...	Independent Designer, Architect or...
54	Taipei	54	Woman	1985-11-17	1985-11-17	Dàtóng/大同	1	Social Medias ; Newspapers	Facebook ; Youtube	Line ; WeChat	Private Sector	Individual Citizen
297	Tallinn	297	Man	1985-12-05	2010-01-01	Kesklinn	3	Social Medias ; Internet	Facebook ; Youtube ; Linkedin ; Inst...	Whatsapp ; Slack	Private Sector	Corporate Company
28	Taipei	28	Woman	1986-01-01	1986-01-01	Zhōngshān/中山	10	Social Medias ; Word to mouth ; Int...	Facebook ; Youtube	Line ; WeChat	Civil Society, Social and Third Sector	Startup & SMEs
281	Tallinn	281	Man	1986-04-03	2010-07-05	Kesklinn	3	Social Medias ; Messaging Apps ;...	Facebook ; Youtube ; Linkedin	Whatsapp ; Messenger but these d...	Private Sector	Corporate Company
158	Tel Aviv	158	Man	1986-04-25	1990-11-28	Ramat... רמת אביב ג'	1	Social Medias ; Television	Facebook ; Youtube ; Linkedin ; Inst...	Whatsapp	Private Sector	Independent Designer, Architect or...

Data Import : Case studies in Python notebook

In [2]:
```
agent = pd.read_sql_query('SELECT * FROM citizen_engagementDB.agent', engine)
agent.head()
```

Out[2]:

	agent_id	city_id	participation_channel	agent_category	agent_type	higly_engaged
0	1	Taipei	Survey	Private Sector	Corporate Company	None
1	2	Taipei	Survey	Public Sector	Individual Citizen	None
2	3	Taipei	Survey	Public Sector	Individual Citizen	None
3	4	Taipei	Survey	Private Sector	Corporate Company	None
4	5	Taipei	Survey	Private Sector	Startup & SMEs	None

In [3]:
```
agent.shape
```

Out[3]:
```
(20989, 6)
```

In [4]:
```
agent.city_id.unique()
```

Out[4]:
```
array(['Taipei', 'Tel Aviv', 'Tallinn'], dtype=object)
```

In [5]:
```
agent.participation_channel.unique()
```

Out[5]:
```
array(['Survey', 'Interview', 'Tweet'], dtype=object)
```

In [6]:
```
agent.agent_type.nunique()
```

Out[6]:
```
12
```

In [7]:
```
agent.agent_category.nunique()
```

Out[7]:
```
5
```

In [8]:
```
tweet = pd.read_sql_query('SELECT * FROM citizen_engagementDB.tweet', engine)
tweet.head()
```

Out[8]:

tweet_id	city_id	agent_id	twitter_profile	tweet_url	content	twe

In [10]:
```
survey = pd.read_sql_query('SELECT * FROM citizen_engagementDB.survey', engine)
survey.head()
```

Out[10]:

survey_id	city_id	agent_id	gender	birth_date	arrival_date	district_id	engagement_feeling	media_source	social_media	..

In [12]:
```
interview = pd.read_sql_query('SELECT * FROM citizen_engagementDB.interview', engine)
interview.head()
```

Out[12]:

interview_id	city_id	agent_id	gender	agent_name	job_position	linkedin_url	organisation	org

In [13]:
```
frames = [agent, survey, interview, tweet]

data = pd.concat(frames,axis=1)
data
```

Out[13]:

	agent_id	city_id	participation_channel	agent_category	agent_type	higly_engaged	survey_id	city_id	agent_id	gender
0	1.0	Taipei	Survey	Private Sector	Corporate Company	None	1.0	Taipei	1.0	Woman
1	2.0	Taipei	Survey	Public Sector	Individual Citizen	None	2.0	Taipei	2.0	Woman

41252 rows × 61 columns

Data Analytics - Cross-Cultural Insights

In [14]:
```
#arrange the data in a decreasing order by the engagement_feeling
#return the top 10 agent_id who feel the most engaged
query = "select agent_id, engagement_feeling \
from citizen_engagementDB.survey \
group by agent_id, engagement_feeling \
order by engagement_feeling desc \
limit 10"
data = pd.read_sql_query(query, engine)
print(data)
#to make it more relevant, I should return the agent_it where engagement_feeling IN ('10')
```

```
   agent_id  engagement_feeling
0       239                  10
1        28                  10
2        44                  10
3       116                  10
4       240                  10
5       270                  10
6        27                  10
7        49                  10
8        84                  10
9       125                  10
```

In [17]:
```
#return the mode value of engagement_feeling
query = "select engagement_feeling, count(engagement_feeling) \
from citizen_engagementDB.survey \
group by engagement_feeling \
having count(engagement_feeling) >= ALL(\
select count(engagement_feeling) \
from citizen_engagementDB.survey \
group by engagement_feeling)"
data = pd.read_sql_query(query, engine)
print(data)
```

```
   engagement_feeling  count(engagement_feeling)
0                   1                        95
```

In [18]:
```
#how many agents per city ?
query = "select city_id, count(agent_id) as total_agents \
from citizen_engagementDB.Agent \
group by city_id \
order by city_id asc"
data = pd.read_sql_query(query, engine)
print(data)
```

```
    city_id  total_agents
0    Taipei          9326
1   Tallinn          3528
2  Tel Aviv          8135
```

In [19]:
```
#how many tweets per twitter_profile and their related city ?
query = "select city_id, twitter_profile, count(twitter_profile) as total_tweets \
from citizen_engagementDB.tweet \
group by city_id, twitter_profile \
order by total_tweets desc"
data = pd.read_sql_query(query, engine)
print(data)
```

```
        city_id                    twitter_profile  total_tweets
0      Tel Aviv       https://twitter.com/Ostrov_A           368
1      Tel Aviv     https://twitter.com/haaretzcom           275
2      Tel Aviv       https://twitter.com/TelAviv            224
3        Taipei  https://twitter.com/CulturalTaiwan           175
4      Tel Aviv  https://twitter.com/Jerusalem_Post           173
...         ...                                ...           ...
21073   Tallinn  https://twitter.com/apocalypticafi             1
21074   Tallinn     https://twitter.com/mjohnsonCoE             1
21075   Tallinn     https://twitter.com/Nimi_naming             1
21076   Tallinn       https://twitter.com/Pardo_Law             1
21077   Tallinn    https://twitter.com/EEAGrants_IN             1

[21078 rows x 3 columns]
```

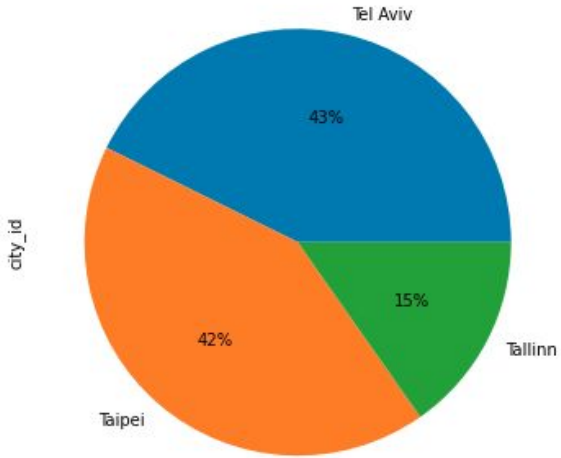
In [25]:
```
#change the default plot size to have a better view of the plots
plot_size = plt.rcParams["figure.figsize"]
print(plot_size[0])
print(plot_size[1])

plot_size[0] = 8
plot_size[1] = 6
plt.rcParams["figure.figsize"] = plot_size
```

```
6.0
4.0
```

In [26]:
```
# plot the number of tweets for each city
tweet.city_id.value_counts().plot(kind='pie', autopct='%1.0f%%')
```

Out[26]:
```
<matplotlib.axes._subplots.AxesSubplot at 0x7fbd3a616d90>
```



Data Visualisation - Tableau

Total survey agents

362

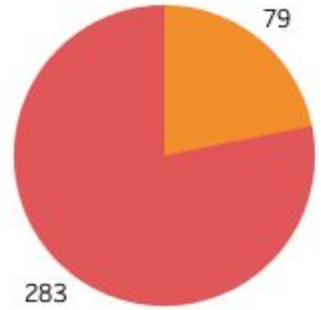
Engag..	
Yes	78,18%
No	21,82%

Engagement Wish

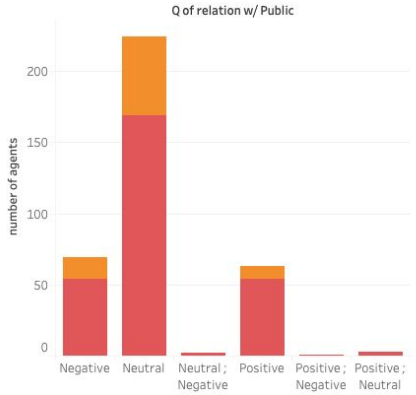
No

Yes

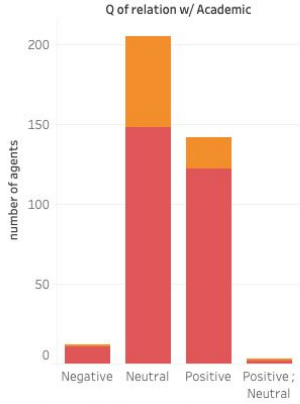
Engagement wish



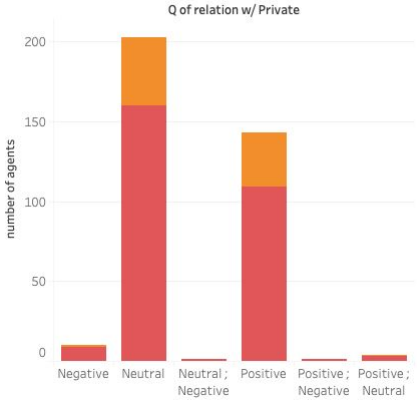
Engagement wish repartition by quality of relationship with Public Sector



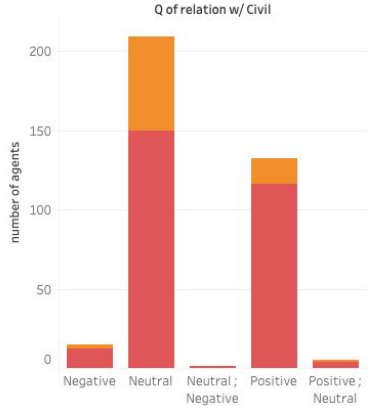
Engagement wish repartition by quality of relationship with Research Sector



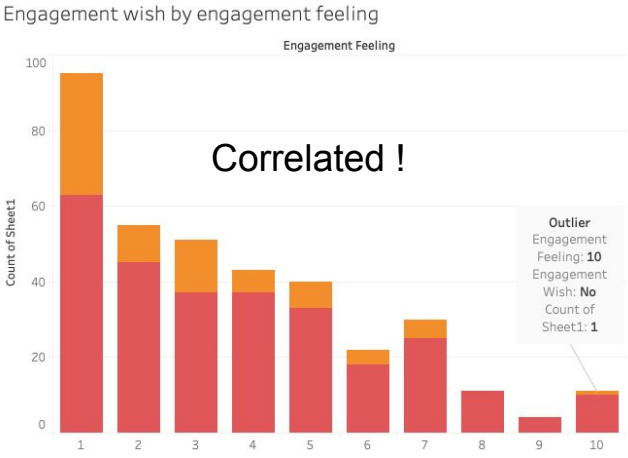
Engagement wish repartition by quality of relationship with Private Sector



Quality of relationship with Civil Sector by engagement wish

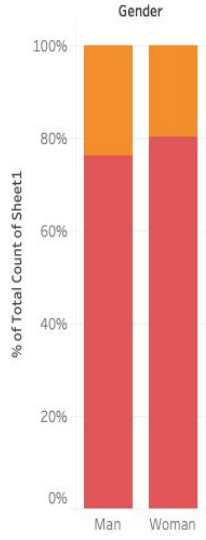


Engagement wish by engagement feeling

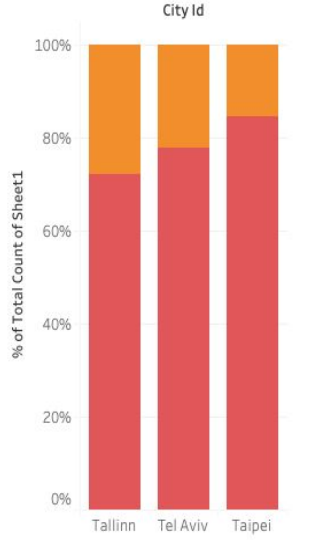


Correlated !

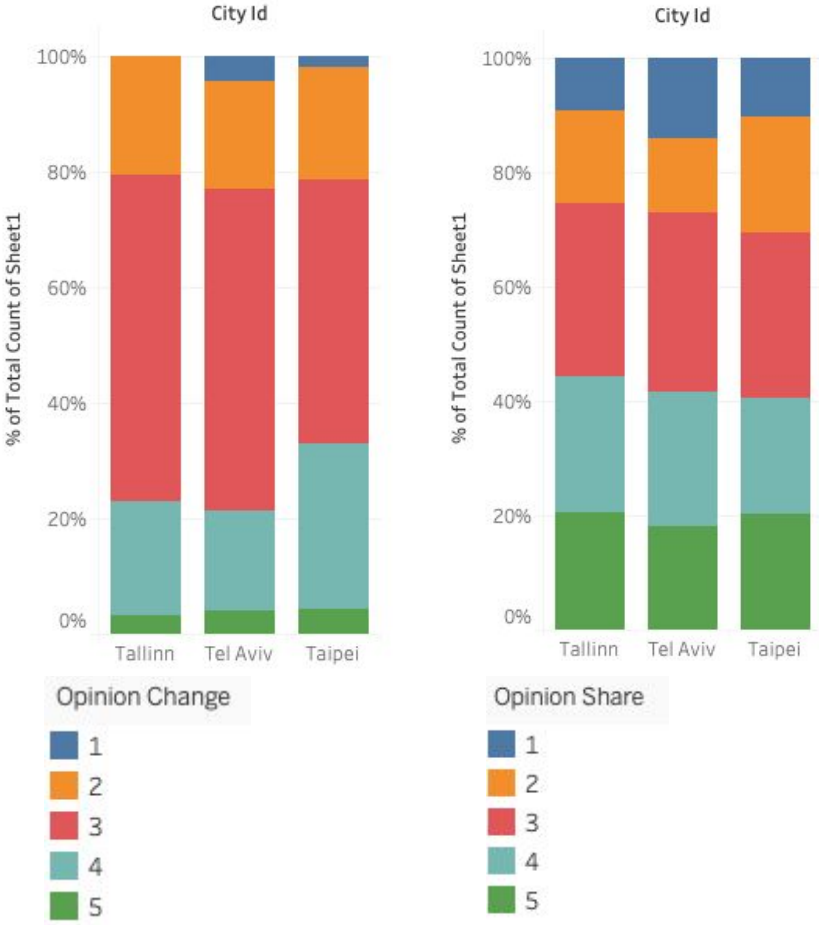
Engagement wish by gender



Engagement wish by city

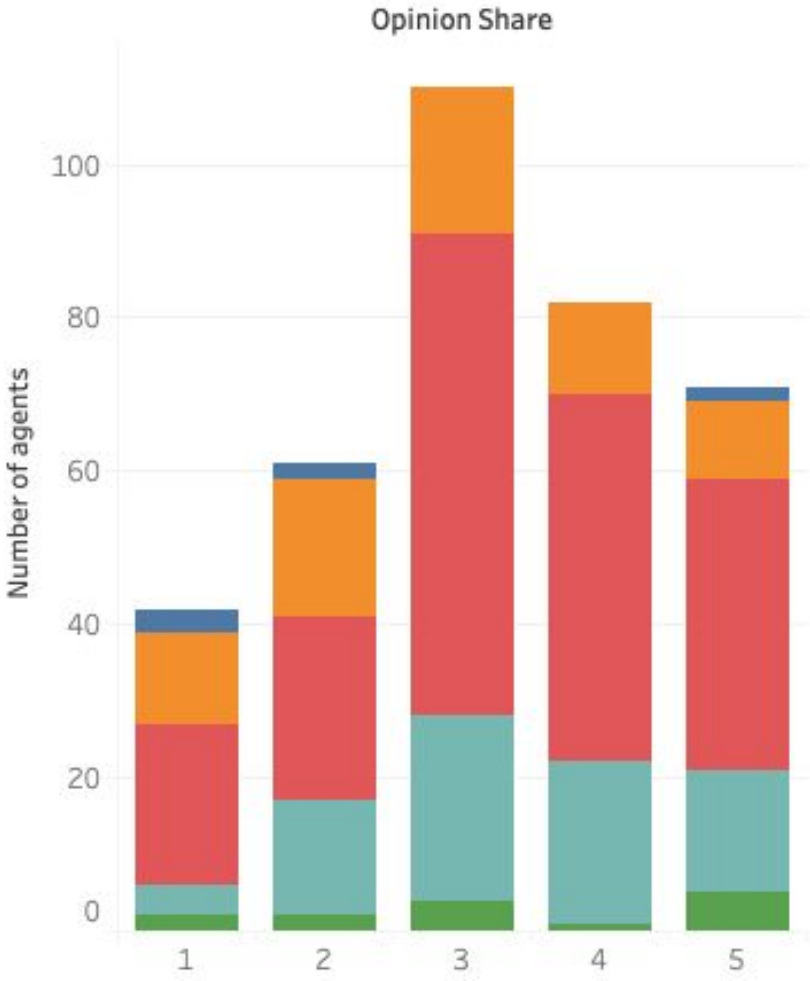


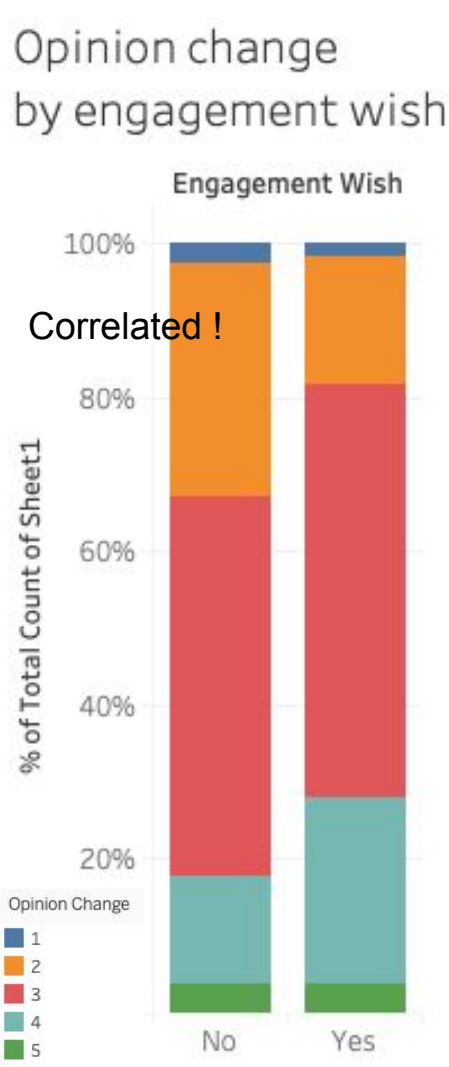
Opinion change by city



Opinion share by city

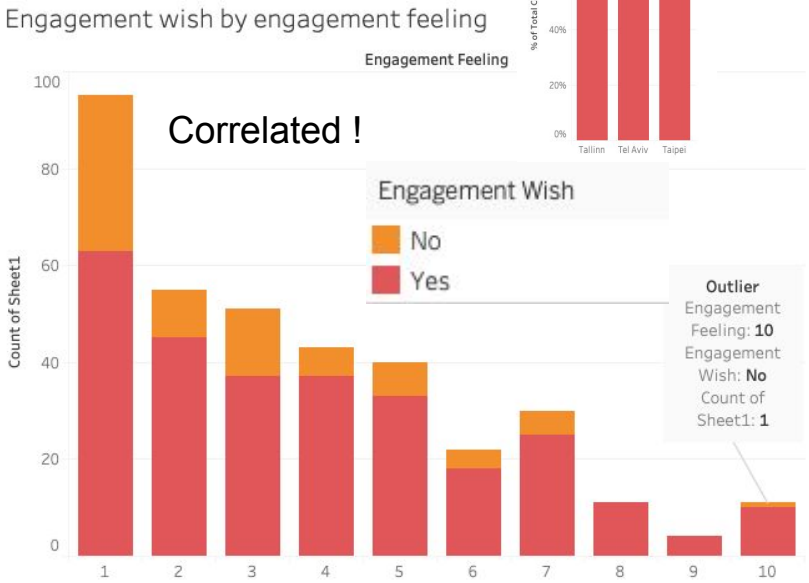
Opinion change level by Opinion share level





Correlated !

Since engagement wish is correlated to engagement feeling, and opinion change is correlated to engagement wish, can we classify cities on these data ? YES



Correlated !