

SQL Scratchpaper and SQL Upload

April 25, 2022

1 Loading

```
[1]: import pandas as pd
from sqlalchemy import create_engine
from sqlalchemy.sql import text
import numpy as np
import datetime
```

```
[2]: engine = create_engine('postgresql://[redacted]:[redacted]@[redacted]:[redacted]/[redacted]')
[redacted]
```

2 Experimentation and Some to_sql

```
[37]: type(datetime.datetime.today() - datetime.timedelta(7))
```

```
[37]: datetime.datetime
```

```
[42]: conn = engine.connect().execution_options(autocommit=True)
```

```
[38]: def query_pm_seven_days(dt):
    start_date = dt - datetime.timedelta(7)
    end_date = dt + datetime.timedelta(7)
    start_SQL = f'SELECT * FROM playrates WHERE "Date" > \'{start_date}\'' AND
    ↪ "Date" <= \'{dt}\';'
    end_SQL = f'SELECT * FROM playrates WHERE "Date" > \'{dt}\'' AND "Date" <=
    ↪ \'{end_date}\';'

    return start_SQL, end_SQL
```

```
[39]: s, e = query_pm_seven_days(datetime.datetime.today() - datetime.timedelta(30))
```

```
[42]: print(e)
```

```
SELECT * FROM playrates WHERE "Date" > '2022-03-22 19:57:10.558436' AND "Date"
<= '2022-03-29 19:57:10.558436';
```

```
[43]: query = conn.execute(text(e))
```

```
[45]: pd.read_sql(s, con=conn)
```

```
[45]:
```

	Date	Champion	Play Rate
0	2022-03-21	Olaf	0.012987
1	2022-03-21	Quinn	0.019980
2	2022-03-21	Maokai	0.015984
3	2022-03-21	Ivern	0.009990
4	2022-03-21	AurelionSol	0.008991
..
472	2022-03-17	Samira	0.083916
473	2022-03-17	Xerath	0.064935
474	2022-03-17	Kayn	0.101898
475	2022-03-17	Singed	0.029970
476	2022-03-17	Ornn	0.024975

[477 rows x 3 columns]

```
[7]: submissions = pd.read_csv("data/df_processing.csv", index_col=[0])
```

```
[67]: df_l_1nf = pd.read_csv('data/df_labels_1nf.csv', index_col=[0])
```

```
[68]: df_l_1nf
```

```
[68]:
```

	submission_id	topic_champion
0	o00m44	twisted fate
1	o00j2e	xerath
2	o00gbv	aatrox
2	o00gbv	ahri
3	o00c4g	aatrox
...
87725	t3tx4z	ahri
87726	t3twyu	aatrox
87726	t3twyu	ahri
87727	t3twhr	aatrox
87727	t3twhr	ahri

[148662 rows x 2 columns]

```
[69]: df_l_1nf.to_sql('submissions_champions', con=conn, index=False)
```

```
[67]: d = int(datetime.datetime(2021,9,15,0,0).timestamp())  
print(d + 1209600)  
print(d - 1209600)
```

1632898800
1630479600

```
[70]: t_table = {'t975':  
    [  
        1000000.0,  
        12.71,  
        4.303,  
        3.182,  
        2.776,  
        2.571,  
        2.447,  
        2.365,  
        2.306,  
        2.262,  
        2.228,  
        2.201,  
        2.179,  
        2.160,  
        2.145,  
        2.131,  
        2.120,  
        2.110,  
        2.101,  
        2.093,  
        2.086,  
        2.080,  
        2.074,  
        2.069,  
        2.064,  
        2.060,  
        2.056,  
        2.052,  
        2.048,  
        2.045,  
        2.042  
    ]  
}
```

```
[74]: pd.DataFrame.from_dict(t_table).to_sql('t_table', con=conn, index=False)
```

```
[9]: submissions.to_sql('submissions_champions_arrays', con=conn, index=False)
```

3 Averaging the Vectors From Word2Vec

```
[3]: comments = pd.read_csv("data/comments_dataframe.csv", index_col=[0])
```

```
[4]: comments
```

```
[4]: submission_id comment_id \
0          o00m44    h1si8vd
1          o00m44    h1sil8e
2          o00m44    h1sigkk
3          o00m44    h1sttyw
4          o00m44    h1sqn3p
...
819317     t3tx4z    hyuijui
819318     t3tx4z    hyujigw
819319     t3tx4z    hyujjc3
819320     t3twyu    hyujgh3
819321     t3twhr    hyuikvu

                                comment_text
0      Roam by having the wave pushed to enemy turret...
1      The first step to roaming is having good wave ...
2      Try not to roam when you think you won't get a...
3                                     Roam less. farm more
4      1 roam kill is one missed wave. One missed wav...
...
819317 Look in event viewer in windows to see what is...
819318 Hi /u/-CrestiaBell. Thank you for participatin...
819319 Im on a similar laptop with W11 and no problem...
819320 Hi /u/DiabloDJ. Thank you for participating in...
819321 Hi /u/NotARussian421. Thank you for participat...

[819322 rows x 3 columns]
```

```
[5]: import gensim
      from tqdm import tqdm
```

```
[6]: model = gensim.models.Word2Vec.load("models/word2vec_5-window_2-min_10-epochs.
      ↪model")
```

```
[7]: avg_vector = list()
      cid_list = list()
```

```
[8]: vocabulary = model.wv.index_to_key
```

```
[9]: for index, row in tqdm(comments.iterrows()):
      comment = row['comment_text']
      cid = row['comment_id']
      tokens = gensim.utils.simple_preprocess(comment)

      valid_vectors = list()

      for token in tokens:
```

```

        if token in vocabulary:
            valid_vectors.append(model.wv[token])
    np_vec = np.mean(valid_vectors, axis=0)
    list_vec = np_vec.tolist()
    cid_list.append(cid)
    avg_vector.append(list_vec)

```

```

470it [00:00, 2126.79it/s]C:\Users\AKost\anaconda3\lib\site-
packages\numpy\core\fromnumeric.py:3419: RuntimeWarning: Mean of empty slice.
    return _methods._mean(a, axis=axis, dtype=dtype,
C:\Users\AKost\anaconda3\lib\site-packages\numpy\core\_methods.py:188:
RuntimeWarning: invalid value encountered in double_scalars
    ret = ret.dtype.type(ret / rcount)
819322it [05:28, 2493.14it/s]

```

```
[10]: r = len(avg_vector)
```

```
[11]: for i in range(r):
        if type(avg_vector[i]) is float:
            avg_vector[i] = [avg_vector[i]]

```

```
[12]: i = 0
        while i < r:
            if len(avg_vector[i]) < 100:
                avg_vector[i] = [None for n in range(100)]
            i += 1

```

```
[ ]: avg_vector
```

```
[14]: len(avg_vector)
```

```
[14]: 819322
```

```
[15]: vdf = pd.DataFrame(avg_vector, index=cid_list)
```

```
[18]: vdf = vdf.reset_index().rename({'index': 'comment_id'}, axis = 'columns')
```

```
[27]: vdf.dtypes
```

```

[27]: comment_id      object
      0             float64
      1             float64
      2             float64
      3             float64
      ...
      95            float64
      96            float64
      97            float64

```

```

98         float64
99         float64
Length: 101, dtype: object

```

```
[32]: temp["comment_id"] = temp["comment_id"].astype(str)
```

```
[33]: comments["comment_id"] = comments["comment_id"].astype(str)
```

```
[21]: temp = comments.drop(columns=['comment_text'])
```

```
[22]: temp
```

```
[22]:
```

	submission_id	comment_id
0	o00m44	h1si8vd
1	o00m44	h1sil8e
2	o00m44	h1sigkk
3	o00m44	h1sttyw
4	o00m44	h1sqn3p
...
819317	t3tx4z	hyuijui
819318	t3tx4z	hyujigw
819319	t3tx4z	hyujjc3
819320	t3twyu	hyujgh3
819321	t3twhr	hyuikvu

[819322 rows x 2 columns]

```
[40]: vdf = vdf.set_index('comment_id').join(temp.set_index('comment_id'),
↳on='comment_id')
```

```
[41]: vdf
```

```
[41]:
```

	0	1	2	3	4	5	\
comment_id							
h1si8vd	-0.049796	-0.541697	-0.395900	0.336320	0.495235	0.079369	
h1sil8e	-0.377742	-0.747743	-0.624726	0.771200	0.106712	-0.407236	
h1sigkk	-0.300410	-0.261556	-0.054437	0.423098	0.178994	0.368531	
h1sttyw	0.915221	-0.156500	1.378459	-1.294745	1.266600	-0.481858	
h1sqn3p	0.070859	0.060563	-0.515558	0.199213	0.220241	0.351475	
...	
hyuijui	-1.355966	0.617499	-0.138140	0.402067	0.208095	-0.014361	
hyujigw	0.110100	1.124392	-0.453732	-0.713067	0.588559	-0.307958	
hyujjc3	-0.221122	0.272347	0.581628	1.032269	-0.563056	-0.015813	
hyujgh3	0.175205	1.205485	-0.472235	-1.251355	0.848424	-0.094128	
hyuikvu	-0.031577	1.231463	-0.578220	-0.666769	0.389533	-0.195280	
	6	7	8	9	...	91	92 \

comment_id					...		
h1si8vd	0.397104	0.619830	0.338848	-0.585261	...	0.817996	0.774166
h1sil8e	-0.094377	0.589819	-0.190857	0.092903	...	0.302844	0.180036
h1sigkk	0.123575	0.223002	0.219818	-0.211503	...	0.283161	0.543719
h1sttyw	0.489934	-0.612194	1.953317	1.429686	...	1.756342	1.854397
h1sqn3p	0.431076	0.459559	0.421736	-0.253068	...	-0.328086	0.731446
...
hyuijui	1.037889	0.308786	0.334837	-0.639551	...	0.937455	-0.352737
hyujigw	-0.574748	0.232228	0.675610	-0.952437	...	0.455434	-0.254409
hyujjc3	0.355627	0.790139	0.085956	-0.906517	...	-0.498260	-0.005669
hyujgh3	-0.337502	0.084641	0.736916	-1.138730	...	0.515949	-0.421447
hyuikvu	-0.716357	0.368137	1.030692	-1.234780	...	0.246419	-0.145363

	93	94	95	96	97	98	\
comment_id							
h1si8vd	-0.434962	0.003623	0.700944	-0.458058	0.254548	1.099897	
h1sil8e	-0.460122	-0.251193	1.274271	-1.185215	0.557564	-0.000997	
h1sigkk	-0.051872	0.176330	0.339582	-0.312815	-0.008801	0.545974	
h1sttyw	-2.366444	0.381295	1.412422	-0.173375	0.736007	1.639724	
h1sqn3p	-0.444711	-0.196509	0.370242	-0.634283	-0.160651	0.816451	
...	
hyuijui	0.428491	-0.626065	-0.145997	-0.086649	-0.166326	0.058679	
hyujigw	-0.566972	-0.059290	0.653864	0.077560	0.649220	-0.508284	
hyujjc3	0.657668	-0.064128	0.130086	-0.239787	-1.286793	0.380034	
hyujgh3	-0.495187	0.123911	0.511195	0.312746	0.989964	-0.560285	
hyuikvu	-0.406440	-0.328594	0.946333	0.352841	0.836917	-0.319618	

	99	submission_id
comment_id		
h1si8vd	-0.318923	o00m44
h1sil8e	-0.353416	o00m44
h1sigkk	-0.755151	o00m44
h1sttyw	-0.102508	o00m44
h1sqn3p	-0.830999	o00m44
...
hyuijui	1.018352	t3tx4z
hyujigw	-0.178613	t3tx4z
hyujjc3	0.603417	t3tx4z
hyujgh3	-0.198824	t3twyu
hyuikvu	0.092641	t3twhr

[819322 rows x 101 columns]

```
[45]: vdf = vdf.reset_index()
```

```
[50]: submission_vectors = vdf.groupby('submission_id').mean()
```

```
[51]: submission_vectors
```

```
[51]:
```

	0	1	2	3	4	5	\
submission_id							
nphcmp	0.427515	0.962019	-0.298306	-0.931229	0.148521	-0.265532	
nphcox	-0.461453	0.348936	0.292223	0.600786	-0.504491	0.381112	
nphdvg	-0.354448	-0.010296	0.290602	0.546764	-0.172587	0.946223	
nphf2u	0.182016	-0.379647	-0.044724	0.525721	-0.194344	-0.459370	
nphg8o	-0.270704	1.065809	-0.033112	-0.029817	0.165749	0.909510	
...	
tsmbvd	-0.561517	-0.034997	0.235754	0.177085	0.188155	0.605975	
tsmc4e	-0.185837	0.397855	-0.265044	-0.009960	0.471385	0.437220	
tsmn3d	-0.413484	1.707641	0.935986	-0.389490	0.486640	0.395572	
tsmp45	-0.698356	-0.154256	0.305001	0.835165	0.116430	0.485211	
tsmr13	0.793250	1.255845	0.082469	1.505851	-2.169148	-0.728317	

	6	7	8	9	...	90	\
submission_id					...		
nphcmp	-0.225392	0.312821	0.351725	-0.826608	...	-0.871377	
nphcox	-0.128308	0.609844	0.154274	-0.316082	...	0.514666	
nphdvg	-0.369047	0.942072	0.419196	-0.244755	...	-0.149547	
nphf2u	-0.452319	0.165828	0.381102	-0.417415	...	0.122480	
nphg8o	0.614613	1.142446	0.164142	-0.793884	...	0.108343	
...	
tsmbvd	0.089307	0.367986	0.160624	-0.814111	...	-0.496271	
tsmc4e	0.093974	-0.124487	0.610498	-0.500805	...	0.184885	
tsmn3d	-0.335198	1.111775	-0.034060	-0.939847	...	-0.648901	
tsmp45	0.803038	0.181127	-0.050343	-0.687846	...	0.309111	
tsmr13	-1.884043	-0.169799	0.108983	-0.010158	...	-1.087428	

	91	92	93	94	95	96	\
submission_id							
nphcmp	0.432817	-0.384758	-0.382450	0.322691	1.208406	0.923407	
nphcox	-0.167961	-0.095978	0.659440	0.053023	0.822105	-0.722880	
nphdvg	0.749959	1.011927	-0.346777	-0.370831	0.012339	-0.357972	
nphf2u	0.141317	0.486795	-0.096966	0.185884	0.107138	-0.534358	
nphg8o	0.430246	0.228479	-0.835066	-0.403983	0.064592	0.819233	
...	
tsmbvd	0.084638	0.732168	0.115819	-0.299866	0.031020	-0.825610	
tsmc4e	-0.014865	0.220821	-0.503964	-0.183408	0.203683	-0.449825	
tsmn3d	1.394263	0.860022	1.819734	0.924424	-0.640734	0.354778	
tsmp45	0.285747	0.849028	-0.225444	-0.721148	-0.115362	-0.625111	
tsmr13	0.719117	0.255539	0.345954	-0.341675	1.201085	0.429511	

	97	98	99
submission_id			
nphcmp	0.520134	-1.038951	-0.081310

nphcox	0.636257	-0.144149	-0.678411
nphdvg	-0.086514	-0.451919	-0.223036
nphf2u	-0.120791	1.181765	-0.590975
nphg8o	-0.138890	-0.340388	-0.272032
...
tsmbvd	-0.834596	0.604650	0.426438
tsmc4e	-0.012341	0.540465	-0.006008
tsmn3d	-0.691425	-0.134965	-0.334619
tsmp45	-0.265894	0.127317	0.767461
tsmr13	-1.135507	-0.561139	0.105068

[87728 rows x 100 columns]

```
[52]: submission_vectors.to_csv('data/submission_vectors.csv')
```

```
[54]: submission_vectors.to_sql('submission_vectors', con=conn, index=True,
    ↳ if_exists='replace')
```