Final_version

December 7, 2017

1 CS 109a Recommendations

1.1 PROJECT INFO

Team Members Maciej Holubiec, Jimena Romero Pinto, Paul von Chamier

---

```
In [5]: import matplotlib.pyplot as plt
        import datetime as dt
        import pandas as pd
        import numpy as np
        import json
        from sklearn.linear_model import Ridge
        from sklearn.linear_model import LassoCV
        from sklearn.model_selection import KFold
        from sklearn.model_selection import GridSearchCV
        from sklearn.linear_model import Lasso
```

2 LOAD DATA

*** Only for the first time. For later uses skip to the "RELOAD DATA" part. Make sure to download smaller (prepocessed) datasets before running it ***

```
In [ ]: # LOAD USER
        df_user = pd.read_json("user.json", lines = True)
```

```
In [ ]: # LOAD BUSINESS
        df_business = pd.read_json("business.json", lines = True)
```

```
In [ ]: # LOAD REVIEWS
        with open('review.json', encoding="utf8") as json_file:
            data_review = json_file.readlines()
            # this line below may take at least 8-10 minutes of processing for 4-5 million rows.
            data_review = list(map(json.loads, data_review))

        df_review = pd.DataFrame(data_review)
```

3 PROCESS

3.0.1 Select restaurants with more than 30 reviews

```
In [ ]: df_business.shape
```

```
In [ ]: df_business = df_business.drop(["hours", "is_open", "latitude", "longitude", "postal_cod
```

```
In [ ]: df_business_350 = df_business[df_business["review_count"] > 350]
```

```
In [ ]: df_business_350.shape
```

3.0.2 Select users who gave more than 100 reviews

```
In [ ]: df_users.shape
```

```
In [ ]: df_user = df_user[["user_id", "review_count"]]
```

```
In [ ]: df_users_150 = df_user[df_user["review_count"] > 150]
```

```
In [ ]: df_users_150.shape
```

3.0.3 Filter out reviews to those corresponding to selected users and restaurants

```
In [ ]: df_review.shape
```

```
In [ ]: df_review = df_review.drop(["cool", "date", "funny", "review_id", "text", "useful"], axi
```

```
In [24]: df_review_350_150 = df_review[df_review["user_id"].isin(df_users_150["user_id"])]
```

```
In [27]: df_review_350_150 = df_review_350_150[df_review_350_150["business_id"].isin(df_business
```

3.0.4 Save data for future reference so we deal with smaller files

```
In [ ]: df_users_150.to_json("df_user_150.json")
```

```
In [ ]: df_business_350.to_json("df_business_350.json")
```

```
In [30]: df_review_350_150.to_json("df_review_350_150.json")
```

4 RELOAD DATA

```
In [82]: df_user = pd.read_json("df_user_100.json")
```

```
In [81]: df_business = pd.read_json("df_business_30.json")
```

```
In [6]: df_review = pd.read_json("df_review_350_150.json")
```

4.0.1 Sample from the data frame because the dataset is still too big

```
In [7]: np.random.seed(9001)
        fraction_of_df = 0.15

In [8]: df_review_smaller = df_review.sample(frac=fraction_of_df)

In [9]: df_review_smaller.shape

Out[9]: (38198, 3)

In [10]: df_review_smaller.to_json("df_review_smaller.json")

In [2]: df_review_smaller = pd.read_json("df_review_smaller.json")
```

4.0.2 Create latent matrix

```
In [11]: r_df = df_review_smaller.pivot(index = 'user_id', columns ='business_id', values = 'sta
         r_df.head()

Out[11]: business_id             --9e1ONYQuAa-CB_Rrw7Tw  -050d_XIor1NpCuWkbIVaQ  \
         user_id
         ---1lKK3aKOuomHnwAkAow                     NaN                     NaN
         --2vR0DIsmQ6WfcSzKWigw                     NaN                     NaN
         --4q8EyqThydQm-eKZpS-A                     NaN                     NaN
         --56mD0sm1eOogphi2FFLw                     NaN                     NaN
         --CIuK7sUpaNzalLAlHJKA                     NaN                     NaN

         business_id             -1xuC540Nycht_iWFeJ-dw  -2ToCaDFpTNmmg3QFzxcWg  \
         user_id
         ---1lKK3aKOuomHnwAkAow                     NaN                     NaN
         --2vR0DIsmQ6WfcSzKWigw                     NaN                     NaN
         --4q8EyqThydQm-eKZpS-A                     NaN                     NaN
         --56mD0sm1eOogphi2FFLw                     NaN                     NaN
         --CIuK7sUpaNzalLAlHJKA                     NaN                     NaN

         business_id             -3zffZUHoY8bQjGfPSoBKQ  -6h3K1hj0d4DRcZNUtHDuw  \
         user_id
         ---1lKK3aKOuomHnwAkAow                     NaN                     NaN
         --2vR0DIsmQ6WfcSzKWigw                     NaN                     NaN
         --4q8EyqThydQm-eKZpS-A                     NaN                     NaN
         --56mD0sm1eOogphi2FFLw                     NaN                     NaN
         --CIuK7sUpaNzalLAlHJKA                     NaN                     NaN

         business_id             -6tvduBzjLI1ISfs3F_qTg  -7H-oXvCxJzuT42ky6Db0g  \
         user_id
         ---1lKK3aKOuomHnwAkAow                     NaN                     NaN
         --2vR0DIsmQ6WfcSzKWigw                     NaN                     NaN
         --4q8EyqThydQm-eKZpS-A                     NaN                     NaN
```

```
--56mD0sm1eOogphi2FFLw                    NaN                    NaN
--CIuK7sUpaNzalLAlHJKA                    NaN                    NaN

business_id            -95mbLJsa0CxXhpaNL4LvA  -9dmhyBvepc08KPEHlEM0w  \
user_id
---1lKK3aKOuomHnwAkAow                    NaN                    NaN
--2vR0DIsmQ6WfcSzKWigw                    NaN                    NaN
--4q8EyqThydQm-eKZpS-A                    NaN                    NaN
--56mD0sm1eOogphi2FFLw                    NaN                    NaN
--CIuK7sUpaNzalLAlHJKA                    NaN                    NaN

business_id                          ...            zcScEL0WEdFkROcnz5379g  \
user_id                              ...
---1lKK3aKOuomHnwAkAow                ...                               NaN
--2vR0DIsmQ6WfcSzKWigw                ...                               NaN
--4q8EyqThydQm-eKZpS-A                ...                               NaN
--56mD0sm1eOogphi2FFLw                ...                               NaN
--CIuK7sUpaNzalLAlHJKA                ...                               NaN

business_id            zdE82PiD6wquvjYLyhOJNA  zgQHtqX0gqMw1nlBZl2VnQ  \
user_id
---1lKK3aKOuomHnwAkAow                    NaN                    NaN
--2vR0DIsmQ6WfcSzKWigw                    NaN                    NaN
--4q8EyqThydQm-eKZpS-A                    NaN                    NaN
--56mD0sm1eOogphi2FFLw                    NaN                    NaN
--CIuK7sUpaNzalLAlHJKA                    NaN                    NaN

business_id            zlpLjbwrKuNs8zROgB_qUQ  znWHLW1pt19HzW1VY6KfCA  \
user_id
---1lKK3aKOuomHnwAkAow                    NaN                    NaN
--2vR0DIsmQ6WfcSzKWigw                    NaN                    NaN
--4q8EyqThydQm-eKZpS-A                    NaN                    NaN
--56mD0sm1eOogphi2FFLw                    NaN                    NaN
--CIuK7sUpaNzalLAlHJKA                    NaN                    NaN

business_id            zoODlH40edpJYLPLkHilNA  zpoZ6WyQUYff18-z4ZU1mA  \
user_id
---1lKK3aKOuomHnwAkAow                    NaN                    NaN
--2vR0DIsmQ6WfcSzKWigw                    NaN                    NaN
--4q8EyqThydQm-eKZpS-A                    NaN                    NaN
--56mD0sm1eOogphi2FFLw                    NaN                    NaN
--CIuK7sUpaNzalLAlHJKA                    NaN                    NaN

business_id            zrDi4gEaUi64lAMfJU51dw  zrTGcb83AsfyVTMrsCa65A  \
user_id
---1lKK3aKOuomHnwAkAow                    NaN                    NaN
--2vR0DIsmQ6WfcSzKWigw                    NaN                    NaN
--4q8EyqThydQm-eKZpS-A                    NaN                    NaN
```

```
--56mDOsm1eOogphi2FFLw                    NaN                    NaN
--CIuK7sUpaNzalLAlHJKA                    NaN                    NaN

business_id             zwNC-Ow4eIMan2__bS9-rg
user_id
---1lKK3aKOuomHnwAkAow                    NaN
--2vRODIsmQ6WfcSzKWigw                    NaN
--4q8EyqThydQm-eKZpS-A                    NaN
--56mDOsm1eOogphi2FFLw                    NaN
--CIuK7sUpaNzalLAlHJKA                    NaN

[5 rows x 1523 columns]
```

```
In [12]: r_df.shape
```

```
Out[12]: (15455, 1523)
```

```
In [13]: fill_zero_rf = r_df.fillna(0)
```

```
In [14]: fill_zero_rf.shape
```

```
Out[14]: (15455, 1523)
```

```
In [15]: fill_zero_rf.head()
```

```
Out[15]: business_id             --9e1ONYQuAa-CB_Rrw7Tw  -050d_XIor1NpCuWkbIVaQ  \
         user_id
         ---1lKK3aKOuomHnwAkAow                     0.0                     0.0
         --2vRODIsmQ6WfcSzKWigw                     0.0                     0.0
         --4q8EyqThydQm-eKZpS-A                     0.0                     0.0
         --56mDOsm1eOogphi2FFLw                     0.0                     0.0
         --CIuK7sUpaNzalLAlHJKA                     0.0                     0.0

         business_id             -1xuC540Nycht_iWFeJ-dw  -2ToCaDFpTNmmg3QFzxcWg  \
         user_id
         ---1lKK3aKOuomHnwAkAow                     0.0                     0.0
         --2vRODIsmQ6WfcSzKWigw                     0.0                     0.0
         --4q8EyqThydQm-eKZpS-A                     0.0                     0.0
         --56mDOsm1eOogphi2FFLw                     0.0                     0.0
         --CIuK7sUpaNzalLAlHJKA                     0.0                     0.0

         business_id             -3zffZUHoY8bQjGfPSoBKQ  -6h3K1hj0d4DRcZNUtHDuw  \
         user_id
         ---1lKK3aKOuomHnwAkAow                     0.0                     0.0
         --2vRODIsmQ6WfcSzKWigw                     0.0                     0.0
         --4q8EyqThydQm-eKZpS-A                     0.0                     0.0
         --56mDOsm1eOogphi2FFLw                     0.0                     0.0
         --CIuK7sUpaNzalLAlHJKA                     0.0                     0.0
```

```
business_id            -6tvduBzjLI1ISfs3F_qTg  -7H-oXvCxJzuT42ky6Db0g  \
user_id
---1lKK3aKOuomHnwAkAow                     0.0                     0.0
--2vR0DIsmQ6WfcSzKWigw                     0.0                     0.0
--4q8EyqThydQm-eKZpS-A                     0.0                     0.0
--56mD0sm1eOogphi2FFLw                     0.0                     0.0
--CIuK7sUpaNzalLAlHJKA                     0.0                     0.0

business_id            -95mbLJsa0CxXhpaNL4LvA  -9dmhyBvepc08KPEHlEM0w  \
user_id
---1lKK3aKOuomHnwAkAow                     0.0                     0.0
--2vR0DIsmQ6WfcSzKWigw                     0.0                     0.0
--4q8EyqThydQm-eKZpS-A                     0.0                     0.0
--56mD0sm1eOogphi2FFLw                     0.0                     0.0
--CIuK7sUpaNzalLAlHJKA                     0.0                     0.0

business_id                         ...           zcScEL0WEdFkROcnz5379g  \
user_id                             ...
---1lKK3aKOuomHnwAkAow              ...                             0.0
--2vR0DIsmQ6WfcSzKWigw              ...                             0.0
--4q8EyqThydQm-eKZpS-A              ...                             0.0
--56mD0sm1eOogphi2FFLw              ...                             0.0
--CIuK7sUpaNzalLAlHJKA              ...                             0.0

business_id            zdE82PiD6wquvjYLyhOJNA  zgQHtqX0gqMw1nlBZl2VnQ  \
user_id
---1lKK3aKOuomHnwAkAow                     0.0                     0.0
--2vR0DIsmQ6WfcSzKWigw                     0.0                     0.0
--4q8EyqThydQm-eKZpS-A                     0.0                     0.0
--56mD0sm1eOogphi2FFLw                     0.0                     0.0
--CIuK7sUpaNzalLAlHJKA                     0.0                     0.0

business_id            zlpLjbwrKuNs8zROgB_qUQ  znWHLW1pt19HzW1VY6KfCA  \
user_id
---1lKK3aKOuomHnwAkAow                     0.0                     0.0
--2vR0DIsmQ6WfcSzKWigw                     0.0                     0.0
--4q8EyqThydQm-eKZpS-A                     0.0                     0.0
--56mD0sm1eOogphi2FFLw                     0.0                     0.0
--CIuK7sUpaNzalLAlHJKA                     0.0                     0.0

business_id            zoODlH40edpJYLPLkHilNA  zpoZ6WyQUYff18-z4ZU1mA  \
user_id
---1lKK3aKOuomHnwAkAow                     0.0                     0.0
--2vR0DIsmQ6WfcSzKWigw                     0.0                     0.0
--4q8EyqThydQm-eKZpS-A                     0.0                     0.0
--56mD0sm1eOogphi2FFLw                     0.0                     0.0
--CIuK7sUpaNzalLAlHJKA                     0.0                     0.0
```

```
business_id             zrDi4gEaUi64lAMfJU51dw  zrTGcb83AsfyVTMrsCa65A  \
user_id
---1lKK3aKOuomHnwAkAow                     0.0                     0.0
--2vR0DIsmQ6WfcSzKWigw                     0.0                     0.0
--4q8EyqThydQm-eKZpS-A                     0.0                     0.0
--56mD0sm1eOogphi2FFLw                     0.0                     0.0
--CIuK7sUpaNzalLAlHJKA                     0.0                     0.0

business_id             zwNC-Ow4eIMan2__bS9-rg
user_id
---1lKK3aKOuomHnwAkAow                     0.0
--2vR0DIsmQ6WfcSzKWigw                     0.0
--4q8EyqThydQm-eKZpS-A                     0.0
--56mD0sm1eOogphi2FFLw                     0.0
--CIuK7sUpaNzalLAlHJKA                     0.0

[5 rows x 1523 columns]
```

5 MODELS

5.0.1 Define RMSE error functions

```
In [16]: def rmse(predictions, targets):
             return np.sqrt(((predictions - targets) ** 2).mean())

In [17]: def rmse2(model, x, y):
             predict = model.predict(x)
             mse2 = rmse(y, predict)
             return mse2
```

5.0.2 Baseline Averages

```
In [18]: avg_mean = r_df.mean().mean()
         avg_mean

Out[18]: 3.8393172911341806

In [19]: rows_length = r_df.shape[0]
         cols_length = r_df.shape[1]

In [20]: cols_means = r_df.mean(axis = 0)
         rows_means = r_df.mean(axis = 1)

In [21]: cols_means.head()

Out[21]: business_id
         --9e1ONYQuAa-CB_Rrw7Tw    4.060606
         -050d_XIor1NpCuWkbIVaQ    3.771429
         -1xuC540Nycht_iWFeJ-dw    4.333333
```

```
-2ToCaDFpTNmmg3QFzxcWg    1.625000
-3zffZUHoY8bQjGfPSoBKQ    4.027778
dtype: float64
```

```
In [22]: rows_means.head()
```

```
Out[22]: user_id
---1lKK3aKOuomHnwAkAow    4.5
--2vR0DIsmQ6WfcSzKWigw    4.5
--4q8EyqThydQm-eKZpS-A    3.0
--56mD0sm1eOogphi2FFLw    4.0
--CIuK7sUpaNzalLAlHJKA    3.0
dtype: float64
```

```
In [23]: preds_array_avg = np.fromfunction(lambda i, j: rows_means[i] + cols_means[j] - avg_mean
```

```
In [24]: preds_array_avg
```

```
Out[24]: array([[ 4.72128877,  4.43211128,  4.99401604, ...,  4.52734938,
                  4.31782557,  4.89145194],
                [ 4.72128877,  4.43211128,  4.99401604, ...,  4.52734938,
                  4.31782557,  4.89145194],
                [ 3.22128877,  2.93211128,  3.49401604, ...,  3.02734938,
                  2.81782557,  3.39145194],
                ...,
                [ 2.72128877,  2.43211128,  2.99401604, ...,  2.52734938,
                  2.31782557,  2.89145194],
                [ 2.22128877,  1.93211128,  2.49401604, ...,  2.02734938,
                  1.81782557,  2.39145194],
                [ 4.22128877,  3.93211128,  4.49401604, ...,  4.02734938,
                  3.81782557,  4.39145194]])
```

```
In [25]: avg_preds_df = pd.DataFrame(preds_array_avg, columns = r_df.columns, index = r_df.index
```

```
In [26]: avg_preds_df.head()
```

```
Out[26]: business_id             --9e1ONYQuAa-CB_Rrw7Tw  -050d_XIor1NpCuWkbIVaQ  \
         user_id
         ---1lKK3aKOuomHnwAkAow                4.721289                4.432111
         --2vR0DIsmQ6WfcSzKWigw                4.721289                4.432111
         --4q8EyqThydQm-eKZpS-A                3.221289                2.932111
         --56mD0sm1eOogphi2FFLw                4.221289                3.932111
         --CIuK7sUpaNzalLAlHJKA                3.221289                2.932111

         business_id             -1xuC540Nycht_iWFeJ-dw  -2ToCaDFpTNmmg3QFzxcWg  \
         user_id
         ---1lKK3aKOuomHnwAkAow                4.994016                2.285683
         --2vR0DIsmQ6WfcSzKWigw                4.994016                2.285683
         --4q8EyqThydQm-eKZpS-A                3.494016                0.785683
```

```
--56mDOsm1eOogphi2FFLw               4.494016               1.785683
--CIuK7sUpaNzalLAlHJKA               3.494016               0.785683

business_id            -3zffZUHoY8bQjGfPSoBKQ  -6h3K1hj0d4DRcZNUtHDuw  \
user_id
---1lKK3aKOuomHnwAkAow               4.68846               3.771794
--2vRODIsmQ6WfcSzKWigw               4.68846               3.771794
--4q8EyqThydQm-eKZpS-A               3.18846               2.271794
--56mDOsm1eOogphi2FFLw               4.18846               3.271794
--CIuK7sUpaNzalLAlHJKA               3.18846               2.271794

business_id            -6tvduBzjLI1ISfs3F_qTg  -7H-oXvCxJzuT42ky6Db0g  \
user_id
---1lKK3aKOuomHnwAkAow               4.271794               4.478865
--2vRODIsmQ6WfcSzKWigw               4.271794               4.478865
--4q8EyqThydQm-eKZpS-A               2.771794               2.978865
--56mDOsm1eOogphi2FFLw               3.771794               3.978865
--CIuK7sUpaNzalLAlHJKA               2.771794               2.978865

business_id            -95mbLJsa0CxXhpaNL4LvA  -9dmhyBvepc08KPEHlEM0w  \
user_id
---1lKK3aKOuomHnwAkAow               4.115228               4.535683
--2vRODIsmQ6WfcSzKWigw               4.115228               4.535683
--4q8EyqThydQm-eKZpS-A               2.615228               3.035683
--56mDOsm1eOogphi2FFLw               3.615228               4.035683
--CIuK7sUpaNzalLAlHJKA               2.615228               3.035683

business_id                             ...            zcScEL0WEdFkROcnz5379g  \
user_id                                 ...
---1lKK3aKOuomHnwAkAow                  ...                          4.374968
--2vRODIsmQ6WfcSzKWigw                  ...                          4.374968
--4q8EyqThydQm-eKZpS-A                  ...                          2.874968
--56mDOsm1eOogphi2FFLw                  ...                          3.874968
--CIuK7sUpaNzalLAlHJKA                  ...                          2.874968

business_id            zdE82PiD6wquvjYLyhOJNA  zgQHtqX0gqMw1nlBZl2VnQ  \
user_id
---1lKK3aKOuomHnwAkAow               4.732111               3.860683
--2vRODIsmQ6WfcSzKWigw               4.732111               3.860683
--4q8EyqThydQm-eKZpS-A               3.232111               2.360683
--56mDOsm1eOogphi2FFLw               4.232111               3.360683
--CIuK7sUpaNzalLAlHJKA               3.232111               2.360683

business_id            zlpLjbwrKuNs8zROgB_qUQ  znWHLW1pt19HzW1VY6KfCA  \
user_id
---1lKK3aKOuomHnwAkAow                3.73963               4.131271
--2vRODIsmQ6WfcSzKWigw                3.73963               4.131271
--4q8EyqThydQm-eKZpS-A                2.23963               2.631271
```

```
--56mDOsm1eOogphi2FFLw                3.23963                3.631271
--CIuK7sUpaNzalLAlHJKA                2.23963                2.631271

business_id             zoODlH4OedpJYLPLkHilNA  zpoZ6WyQUYff18-z4ZU1mA  \
user_id
---1lKK3aKOuomHnwAkAow               5.182422                4.994016
--2vRODIsmQ6WfcSzKWigw               5.182422                4.994016
--4q8EyqThydQm-eKZpS-A               3.682422                3.494016
--56mDOsm1eOogphi2FFLw               4.682422                4.494016
--CIuK7sUpaNzalLAlHJKA               3.682422                3.494016

business_id             zrDi4gEaUi64lAMfJU51dw  zrTGcb83AsfyVTMrsCa65A  \
user_id
---1lKK3aKOuomHnwAkAow               4.527349                4.317826
--2vRODIsmQ6WfcSzKWigw               4.527349                4.317826
--4q8EyqThydQm-eKZpS-A               3.027349                2.817826
--56mDOsm1eOogphi2FFLw               4.027349                3.817826
--CIuK7sUpaNzalLAlHJKA               3.027349                2.817826

business_id             zwNC-Ow4eIMan2__bS9-rg
user_id
---1lKK3aKOuomHnwAkAow               4.891452
--2vRODIsmQ6WfcSzKWigw               4.891452
--4q8EyqThydQm-eKZpS-A               3.391452
--56mDOsm1eOogphi2FFLw               4.391452
--CIuK7sUpaNzalLAlHJKA               3.391452

[5 rows x 1523 columns]
```

```
In [107]: rmse_pd = pd.DataFrame()
          rmse_pd = (avg_preds_df - r_df)**2
          avg_mse = rmse_pd.mean().mean()
          avg_rmse = np.sqrt(avg_mse)
```

```
In [108]: avg_rmse
```

```
Out[108]: 0.7595952486501506
```

5.0.3 Baseline Regression

```
In [29]: categorical_columns = ['business_id', 'user_id']
```

```
In [30]: unique_business = df_review_smaller.business_id.nunique()
         unique_business
```

```
Out[30]: 1523
```

```
In [31]: unique_user = df_review_smaller.user_id.nunique()
         unique_user
```

```
Out[31]: 15455

In [32]: df_review_dummies = pd.get_dummies(df_review_smaller, columns=categorical_columns, drop

In [33]: df_review_dummies.shape

Out[33]: (38198, 16979)

In [34]: # df_review_dummies.to_json("df_review_dummies.json")

In [35]: # df_review_smaller = pd.read_json("df_review_smaller.json")

In [36]: np.random.seed(9001)
         msk = np.random.rand(len(df_review_dummies)) < 0.5

         # data_train = df_subset[msk]
         # data_test = df_subset[~msk]

         x_train = df_review_dummies[msk].drop(['stars'], axis=1) # DataFrame

         x_test = df_review_dummies[~msk].drop(['stars'], axis=1) # DataFrame

         y_train = df_review_dummies[msk].stars #series

         y_test = df_review_dummies[~msk].stars # series

In [37]: ols_lasso = Lasso(alpha=0.0001)
         ols_lasso.fit(x_train,y_train)

Out[37]: Lasso(alpha=0.0001, copy_X=True, fit_intercept=True, max_iter=1000,
            normalize=False, positive=False, precompute=False, random_state=None,
            selection='cyclic', tol=0.0001, warm_start=False)

In [38]: rmse2(ols_lasso, x_train, y_train)

Out[38]: 0.86079140172462165

In [39]: rmse2(ols_lasso, x_test, y_test)

Out[39]: 0.97445307396512448

In [40]: y_preds = ols_lasso.predict(x_train)

In [41]: ols_lasso.coef_

Out[41]: array([ 0.29526438,  0.        ,  0.08687282, ..., -0.31021108,
                -0.        , -0.        ])

In [42]: busienss_coeffs = ols_lasso.coef_[:unique_business]

In [43]: user_coeffs = ols_lasso.coef_[unique_business:]
```

```
In [44]: user_coeffs.shape, busienss_coeffs.shape

Out[44]: ((15455,), (1523,))

In [45]: preds_array_reg = np.fromfunction(lambda i, j: user_coeffs[i] + busienss_coeffs[j] + av

In [46]: preds_array_reg

Out[46]: array([[ 4.13458167,  3.83931729,  3.92619012, ...,  3.83931729,
                  3.71403969,  3.83931729],
                [ 4.13458167,  3.83931729,  3.92619012, ...,  3.83931729,
                  3.71403969,  3.83931729],
                [ 4.13458167,  3.83931729,  3.92619012, ...,  3.83931729,
                  3.71403969,  3.83931729],
                ...,
                [ 3.82437059,  3.52910621,  3.61597903, ...,  3.52910621,
                  3.40382861,  3.52910621],
                [ 4.13458167,  3.83931729,  3.92619012, ...,  3.83931729,
                  3.71403969,  3.83931729],
                [ 4.13458167,  3.83931729,  3.92619012, ...,  3.83931729,
                  3.71403969,  3.83931729]])

In [47]: preds_array_reg_df = pd.DataFrame(preds_array_reg, columns = r_df.columns, index = r_df

In [48]: preds_array_reg_df.head()

Out[48]: business_id             --9e1ONYQuAa-CB_Rrw7Tw  -050d_XIor1NpCuWkbIVaQ  \
         user_id
         ---1lKK3aKOuomHnwAkAow               4.134582                3.839317
         --2vR0DIsmQ6WfcSzKWigw               4.134582                3.839317
         --4q8EyqThydQm-eKZpS-A               4.134582                3.839317
         --56mD0sm1eOogphi2FFLw               4.134582                3.839317
         --CIuK7sUpaNzalLAlHJKA               4.134582                3.839317

         business_id             -1xuC540Nycht_iWFeJ-dw  -2ToCaDFpTNmmg3QFzxcWg  \
         user_id
         ---1lKK3aKOuomHnwAkAow                 3.92619                1.540696
         --2vR0DIsmQ6WfcSzKWigw                 3.92619                1.540696
         --4q8EyqThydQm-eKZpS-A                 3.92619                1.540696
         --56mD0sm1eOogphi2FFLw                 3.92619                1.540696
         --CIuK7sUpaNzalLAlHJKA                 3.92619                1.540696

         business_id             -3zffZUHoY8bQjGfPSoBKQ  -6h3K1hj0d4DRcZNUtHDuw  \
         user_id
         ---1lKK3aKOuomHnwAkAow               3.928585                2.874029
         --2vR0DIsmQ6WfcSzKWigw               3.928585                2.874029
         --4q8EyqThydQm-eKZpS-A               3.928585                2.874029
         --56mD0sm1eOogphi2FFLw               3.928585                2.874029
         --CIuK7sUpaNzalLAlHJKA               3.928585                2.874029
```

```
business_id             -6tvduBzjLI1ISfs3F_qTg  -7H-oXvCxJzuT42ky6Db0g  \
user_id
---1lKK3aKOuomHnwAkAow                3.839317                3.839317
--2vR0DIsmQ6WfcSzKWigw                3.839317                3.839317
--4q8EyqThydQm-eKZpS-A                3.839317                3.839317
--56mD0sm1eOogphi2FFLw                3.839317                3.839317
--CIuK7sUpaNzalLAlHJKA                3.839317                3.839317

business_id             -95mbLJsa0CxXhpaNL4LvA  -9dmhyBvepc08KPEHlEM0w  \
user_id
---1lKK3aKOuomHnwAkAow                3.839317                3.760319
--2vR0DIsmQ6WfcSzKWigw                3.839317                3.760319
--4q8EyqThydQm-eKZpS-A                3.839317                3.760319
--56mD0sm1eOogphi2FFLw                3.839317                3.760319
--CIuK7sUpaNzalLAlHJKA                3.839317                3.760319

business_id                      ...            zcScEL0WEdFkROcnz5379g  \
user_id                          ...
---1lKK3aKOuomHnwAkAow           ...                          3.574515
--2vR0DIsmQ6WfcSzKWigw           ...                          3.574515
--4q8EyqThydQm-eKZpS-A           ...                          3.574515
--56mD0sm1eOogphi2FFLw           ...                          3.574515
--CIuK7sUpaNzalLAlHJKA           ...                          3.574515

business_id             zdE82PiD6wquvjYLyhOJNA  zgQHtqX0gqMw1nlBZl2VnQ  \
user_id
---1lKK3aKOuomHnwAkAow                3.904549                3.254626
--2vR0DIsmQ6WfcSzKWigw                3.904549                3.254626
--4q8EyqThydQm-eKZpS-A                3.904549                3.254626
--56mD0sm1eOogphi2FFLw                3.904549                3.254626
--CIuK7sUpaNzalLAlHJKA                3.904549                3.254626

business_id             zlpLjbwrKuNs8zROgB_qUQ  znWHLW1pt19HzW1VY6KfCA  \
user_id
---1lKK3aKOuomHnwAkAow                2.938128                3.766569
--2vR0DIsmQ6WfcSzKWigw                2.938128                3.766569
--4q8EyqThydQm-eKZpS-A                2.938128                3.766569
--56mD0sm1eOogphi2FFLw                2.938128                3.766569
--CIuK7sUpaNzalLAlHJKA                2.938128                3.766569

business_id             zoODlH40edpJYLPLkHilNA  zpoZ6WyQUYff18-z4ZU1mA  \
user_id
---1lKK3aKOuomHnwAkAow                3.977102                3.925678
--2vR0DIsmQ6WfcSzKWigw                3.977102                3.925678
--4q8EyqThydQm-eKZpS-A                3.977102                3.925678
--56mD0sm1eOogphi2FFLw                3.977102                3.925678
--CIuK7sUpaNzalLAlHJKA                3.977102                3.925678
```

```
business_id             zrDi4gEaUi64lAMfJU51dw  zrTGcb83AsfyVTMrsCa65A  \
user_id
---1lKK3aKOuomHnwAkAow                3.839317                 3.71404
--2vR0DIsmQ6WfcSzKWigw                3.839317                 3.71404
--4q8EyqThydQm-eKZpS-A                3.839317                 3.71404
--56mD0sm1eOogphi2FFLw                3.839317                 3.71404
--CIuK7sUpaNzalLAlHJKA                3.839317                 3.71404

business_id             zwNC-Ow4eIMan2__bS9-rg
user_id
---1lKK3aKOuomHnwAkAow                3.839317
--2vR0DIsmQ6WfcSzKWigw                3.839317
--4q8EyqThydQm-eKZpS-A                3.839317
--56mD0sm1eOogphi2FFLw                3.839317
--CIuK7sUpaNzalLAlHJKA                3.839317

[5 rows x 1523 columns]
```

In [53]:
```
resids_array = np.subtract(r_df, preds_array_reg)
```

In [58]:
```
resids_df = resids_array.fillna(0)
```

In [59]:
```
resids_df.head()
```

Out[59]:
```
business_id             --9e1ONYQuAa-CB_Rrw7Tw  -050d_XIor1NpCuWkbIVaQ  \
user_id
---1lKK3aKOuomHnwAkAow                     0.0                     0.0
--2vR0DIsmQ6WfcSzKWigw                     0.0                     0.0
--4q8EyqThydQm-eKZpS-A                     0.0                     0.0
--56mD0sm1eOogphi2FFLw                     0.0                     0.0
--CIuK7sUpaNzalLAlHJKA                     0.0                     0.0

business_id             -1xuC540Nycht_iWFeJ-dw  -2ToCaDFpTNmmg3QFzxcWg  \
user_id
---1lKK3aKOuomHnwAkAow                     0.0                     0.0
--2vR0DIsmQ6WfcSzKWigw                     0.0                     0.0
--4q8EyqThydQm-eKZpS-A                     0.0                     0.0
--56mD0sm1eOogphi2FFLw                     0.0                     0.0
--CIuK7sUpaNzalLAlHJKA                     0.0                     0.0

business_id             -3zffZUHoY8bQjGfPSoBKQ  -6h3K1hj0d4DRcZNUtHDuw  \
user_id
---1lKK3aKOuomHnwAkAow                     0.0                     0.0
--2vR0DIsmQ6WfcSzKWigw                     0.0                     0.0
--4q8EyqThydQm-eKZpS-A                     0.0                     0.0
--56mD0sm1eOogphi2FFLw                     0.0                     0.0
--CIuK7sUpaNzalLAlHJKA                     0.0                     0.0
```

```
business_id             -6tvduBzjLI1ISfs3F_qTg  -7H-oXvCxJzuT42ky6Db0g  \
user_id
---1lKK3aKOuomHnwAkAow                     0.0                     0.0
--2vR0DIsmQ6WfcSzKWigw                     0.0                     0.0
--4q8EyqThydQm-eKZpS-A                     0.0                     0.0
--56mD0sm1eOogphi2FFLw                     0.0                     0.0
--CIuK7sUpaNzalLAlHJKA                     0.0                     0.0

business_id             -95mbLJsa0CxXhpaNL4LvA  -9dmhyBvepc08KPEHlEM0w  \
user_id
---1lKK3aKOuomHnwAkAow                     0.0                     0.0
--2vR0DIsmQ6WfcSzKWigw                     0.0                     0.0
--4q8EyqThydQm-eKZpS-A                     0.0                     0.0
--56mD0sm1eOogphi2FFLw                     0.0                     0.0
--CIuK7sUpaNzalLAlHJKA                     0.0                     0.0

business_id                        ...          zcScEL0WEdFkROcnz5379g  \
user_id                            ...
---1lKK3aKOuomHnwAkAow             ...                             0.0
--2vR0DIsmQ6WfcSzKWigw             ...                             0.0
--4q8EyqThydQm-eKZpS-A             ...                             0.0
--56mD0sm1eOogphi2FFLw             ...                             0.0
--CIuK7sUpaNzalLAlHJKA             ...                             0.0

business_id             zdE82PiD6wquvjYLyhOJNA  zgQHtqX0gqMw1nlBZl2VnQ  \
user_id
---1lKK3aKOuomHnwAkAow                     0.0                     0.0
--2vR0DIsmQ6WfcSzKWigw                     0.0                     0.0
--4q8EyqThydQm-eKZpS-A                     0.0                     0.0
--56mD0sm1eOogphi2FFLw                     0.0                     0.0
--CIuK7sUpaNzalLAlHJKA                     0.0                     0.0

business_id             zlpLjbwrKuNs8zROgB_qUQ  znWHLW1pt19HzW1VY6KfCA  \
user_id
---1lKK3aKOuomHnwAkAow                     0.0                     0.0
--2vR0DIsmQ6WfcSzKWigw                     0.0                     0.0
--4q8EyqThydQm-eKZpS-A                     0.0                     0.0
--56mD0sm1eOogphi2FFLw                     0.0                     0.0
--CIuK7sUpaNzalLAlHJKA                     0.0                     0.0

business_id             zoODlH40edpJYLPLkHilNA  zpoZ6WyQUYff18-z4ZU1mA  \
user_id
---1lKK3aKOuomHnwAkAow                     0.0                     0.0
--2vR0DIsmQ6WfcSzKWigw                     0.0                     0.0
--4q8EyqThydQm-eKZpS-A                     0.0                     0.0
--56mD0sm1eOogphi2FFLw                     0.0                     0.0
--CIuK7sUpaNzalLAlHJKA                     0.0                     0.0
```

```
business_id             zrDi4gEaUi64lAMfJU51dw  zrTGcb83AsfyVTMrsCa65A  \
user_id
---1lKK3aKOuomHnwAkAow                     0.0                     0.0
--2vR0DIsmQ6WfcSzKWigw                     0.0                     0.0
--4q8EyqThydQm-eKZpS-A                     0.0                     0.0
--56mD0sm1eOogphi2FFLw                     0.0                     0.0
--CIuK7sUpaNzalLAlHJKA                     0.0                     0.0

business_id             zwNC-Ow4eIMan2__bS9-rg
user_id
---1lKK3aKOuomHnwAkAow                     0.0
--2vR0DIsmQ6WfcSzKWigw                     0.0
--4q8EyqThydQm-eKZpS-A                     0.0
--56mD0sm1eOogphi2FFLw                     0.0
--CIuK7sUpaNzalLAlHJKA                     0.0

[5 rows x 1523 columns]
```

```
In [114]: rmse_pd = pd.DataFrame()
          rmse_pd = (preds_array_reg_df - r_df)**2
          reg_mse = rmse_pd.mean().mean()
          reg_rmse = np.sqrt(reg_mse)
```

```
In [115]: reg_rmse
```

```
Out[115]: 0.93543616311236566
```

5.0.4 Matrix Factorization

Code taken from https://bugra.github.io/work/notes/2014-04-19/alternating-least-squares-method-for-collaborative-filtering/

```
In [61]: Q = resids_df.values
```

```
In [62]: W = Q>0.5
         W[W == True] = 1
         W[W == False] = 0
         # To be consistent with our Q matrix
         W = W.astype(np.float64, copy=False)
```

```
In [63]: W.shape
```

```
Out[63]: (15455, 1523)
```

```
In [ ]: def get_error(Q, X, Y, W):
            return np.sum((W * (Q - np.dot(X, Y)))**2)
```

We do not really know how to select data (I discussed it with the TF during office hours) for train/valid/test on this model so we are performing parameter choice on the entire sample that we have (which might work as a train set, with test set being all the users/restaurants that we skipped when selecting a small sample size that will work). Parameter choice is below and shows that we should pick 100 latent factors and lambda of 0.1:

In [124]:
```
lambda_ = [0.01, 0.1, 1, 10, 100]
n_fac = [5, 10, 20, 50, 100]
m, n = Q.shape
n_iterations = 10
```

In [126]:
```
error_pair = {}
for l in lambda_:
    for n_factors in n_fac:
        X = 5 * np.random.rand(m, n_factors)
        Y = 5 * np.random.rand(n_factors, n)
        errors = []
        for ii in range(n_iterations):
            X = np.linalg.solve(np.dot(Y, Y.T) + l * np.eye(n_factors),
                                np.dot(Y, Q.T)).T
            Y = np.linalg.solve(np.dot(X.T, X) + l * np.eye(n_factors),
                                np.dot(X.T, Q))
            errors.append(get_error(Q, X, Y, W))
        error_current = errors[-1]
        print(l,n_factors, error_current)
        error_pair[(l,n_factors)] = error_current
```

```
0.01 5 14293.1539642
0.01 10 13760.3267322
0.01 20 12970.6802666
0.01 50 11392.4976871
0.01 100 9589.84479649
0.1 5 14272.5471682
0.1 10 13742.9858822
0.1 20 12941.7621586
0.1 50 11342.4390983
0.1 100 9565.59825812
1 5 14277.8991661
1 10 13778.7463084
1 20 13001.5147731
1 50 11424.2345257
1 100 9707.7889593
10 5 14583.8993997
10 10 14407.4591845
10 20 14230.9229841
10 50 14119.4566875
10 100 14065.0777633
100 5 14893.2815749
100 10 14893.2815749
100 20 14893.2815749
100 50 14893.2815749
100 100 14893.2815749
```

In [127]:
```
lambda_ = 0.1
```

```
n_factors = 100
m, n = Q.shape
n_iterations = 10
```

In [128]:
```
X = 5 * np.random.rand(m, n_factors)
Y = 5 * np.random.rand(n_factors, n)
```

In [129]:
```
errors = []
for ii in range(n_iterations):
    X = np.linalg.solve(np.dot(Y, Y.T) + lambda_ * np.eye(n_factors),
                        np.dot(Y, Q.T)).T
    Y = np.linalg.solve(np.dot(X.T, X) + lambda_ * np.eye(n_factors),
                        np.dot(X.T, Q))
    errors.append(get_error(Q, X, Y, W))
Q_hat = np.dot(X, Y)
print('Error of rated movies: {}'.format(get_error(Q, X, Y, W)))
```

```
Error of rated movies: 9564.656805985478
```

In [130]:
```
fac_preds_df = pd.DataFrame(Q_hat, columns = r_df.columns, index = r_df.index)
```

In [131]:
```
fac_preds_df.head()
```

Out[131]:

business_id	--9e1ONYQuAa-CB_Rrw7Tw	-050d_XIor1NpCuWkbIVaQ	\
user_id			
---1lKK3aKOuomHnwAkAow	0.013486	0.001198	
--2vR0DIsmQ6WfcSzKWigw	-0.001610	0.000798	
--4q8EyqThydQm-eKZpS-A	0.010395	0.000422	
--56mD0sm1eOogphi2FFLw	0.000135	-0.000032	
--CIuK7sUpaNzalLAlHJKA	-0.024051	0.005735	

business_id	-1xuC540Nycht_iWFeJ-dw	-2ToCaDFpTNmmg3QFzxcWg	\
user_id			
---1lKK3aKOuomHnwAkAow	-7.542290e-05	-0.000041	
--2vR0DIsmQ6WfcSzKWigw	-2.078653e-04	-0.000194	
--4q8EyqThydQm-eKZpS-A	3.098898e-04	0.000061	
--56mD0sm1eOogphi2FFLw	-2.571468e-07	-0.000012	
--CIuK7sUpaNzalLAlHJKA	3.490087e-04	0.000094	

business_id	-3zffZUHoY8bQjGfPSoBKQ	-6h3K1hj0d4DRcZNUtHDuw	\
user_id			
---1lKK3aKOuomHnwAkAow	-0.012063	-6.264886e-03	
--2vR0DIsmQ6WfcSzKWigw	0.006069	9.506521e-05	
--4q8EyqThydQm-eKZpS-A	0.001071	1.852691e-04	
--56mD0sm1eOogphi2FFLw	-0.000109	-6.915962e-07	
--CIuK7sUpaNzalLAlHJKA	-0.001931	-4.595633e-04	

business_id	-6tvduBzjLI1ISfs3F_qTg	-7H-oXvCxJzuT42ky6Db0g	\

```
user_id
---1lKK3aKOuomHnwAkAow              -0.000405           1.936014e-05
--2vR0DIsmQ6WfcSzKWigw               0.003577          -3.329624e-05
--4q8EyqThydQm-eKZpS-A              -0.000589           8.232104e-06
--56mD0sm1eOogphi2FFLw               0.000149          -3.894938e-08
--CIuK7sUpaNzalLAlHJKA               0.000034           1.590084e-06

business_id            -95mbLJsa0CxXhpaNL4LvA  -9dmhyBvepc08KPEHlEM0w  \
user_id
---1lKK3aKOuomHnwAkAow             -0.000117                0.000808
--2vR0DIsmQ6WfcSzKWigw              0.000273                0.000293
--4q8EyqThydQm-eKZpS-A              0.000088                0.000490
--56mD0sm1eOogphi2FFLw             -0.000049                0.000009
--CIuK7sUpaNzalLAlHJKA             -0.005689                0.001583

business_id                  ...           zcScEL0WEdFkROcnz5379g  \
user_id                      ...
---1lKK3aKOuomHnwAkAow       ...                         0.000886
--2vR0DIsmQ6WfcSzKWigw       ...                        -0.001108
--4q8EyqThydQm-eKZpS-A       ...                        -0.000579
--56mD0sm1eOogphi2FFLw       ...                         0.000071
--CIuK7sUpaNzalLAlHJKA       ...                         0.003711

business_id            zdE82PiD6wquvjYLyhOJNA  zgQHtqX0gqMw1nlBZl2VnQ  \
user_id
---1lKK3aKOuomHnwAkAow              0.009329                0.000468
--2vR0DIsmQ6WfcSzKWigw              0.000924                0.000384
--4q8EyqThydQm-eKZpS-A             -0.003090                -0.000264
--56mD0sm1eOogphi2FFLw              0.000217                0.000006
--CIuK7sUpaNzalLAlHJKA              0.000371                0.000499

business_id            zlpLjbwrKuNs8zROgB_qUQ  znWHLW1pt19HzW1VY6KfCA  \
user_id
---1lKK3aKOuomHnwAkAow              0.004172                -0.000235
--2vR0DIsmQ6WfcSzKWigw              0.001234                0.001395
--4q8EyqThydQm-eKZpS-A              0.003278                -0.004647
--56mD0sm1eOogphi2FFLw              0.000089                0.000146
--CIuK7sUpaNzalLAlHJKA             -0.023474                -0.006257

business_id            zoODlH40edpJYLPLkHilNA  zpoZ6WyQUYff18-z4ZU1mA  \
user_id
---1lKK3aKOuomHnwAkAow              0.003479                0.001407
--2vR0DIsmQ6WfcSzKWigw             -0.009670                -0.005757
--4q8EyqThydQm-eKZpS-A              0.000186                -0.000196
--56mD0sm1eOogphi2FFLw             -0.000165                -0.000024
--CIuK7sUpaNzalLAlHJKA             -0.015964                0.002198

business_id            zrDi4gEaUi64lAMfJU51dw  zrTGcb83AsfyVTMrsCa65A  \
```

```
user_id
---1lKK3aKOuomHnwAkAow               -0.000440               0.000193
--2vR0DIsmQ6WfcSzKWigw               -0.001256               0.000372
--4q8EyqThydQm-eKZpS-A                0.000293               0.000236
--56mD0sm1eOogphi2FFLw               -0.000062              -0.000022
--CIuK7sUpaNzalLAlHJKA                0.000598               0.002265

business_id             zwNC-Ow4eIMan2__bS9-rg
user_id
---1lKK3aKOuomHnwAkAow               -0.009090
--2vR0DIsmQ6WfcSzKWigw               -0.002269
--4q8EyqThydQm-eKZpS-A                0.000214
--56mD0sm1eOogphi2FFLw                0.000010
--CIuK7sUpaNzalLAlHJKA               -0.003075

[5 rows x 1523 columns]
```

```
In [132]: rmse_pd = pd.DataFrame()
          rmse_pd = (fac_preds_df + preds_array_reg_df - r_df)**2
          fac_mse = rmse_pd.mean().mean()
          fac_rmse = np.sqrt(fac_mse)
```

```
In [133]: fac_rmse
```

```
Out[133]: 0.87682149384685815
```

6 PREDICT

```
In [83]: def get_reccomendations(user, number_rec,df):
             top_preds = df.loc[user][fill_zero_rf.loc[user] == 0].sort_values(ascending = False
             top_preds_df = pd.DataFrame(top_preds).rename(columns={user:"predicted rating"})
             predictions = pd.merge(left = top_preds_df, right = df_business, left_index = True,
             return predictions
```

Let's predict for user ---1lKK3aKOuomHnwAkAow. Let's check what are his top choices:

```
In [105]: top_ratings_user_x = df_review[df_review["user_id"] == "---1lKK3aKOuomHnwAkAow"].sort_
```

```
In [106]: df_business[df_business["business_id"].isin(top_ratings_user_x)]
```

```
Out[106]:                                              address  \
          104700                      750 S Rampart Blvd, Ste 7
          110934                                  113 N 4th St
          142630                3555 S Town Center Dr, Ste 105
          14551                       750 S Rampart Blvd, Ste 9
          32230                             440 S Rampart Blvd
          40479              10100 W Charleston Blvd, Ste 150
          78134                             8975 S Eastern Ave
```

```
84520   The Mirage Hotel Casino, 3400 Las Vegas Blvd S
92918                          953 E Sahara Ave, Ste A5
93528                  8751 W Charleston Blvd, Ste 110

                  business_id  \
104700  RRw9I8pHt5PzgYGT2QeODw
110934  eJKnymd0BywNPrJw1IuXVw
142630  bPcqucuuClxYrIM8xWoArg
14551   rq5dgoksPHkJwJNQKlGQ7w
32230   igHYkXZMLAc9UdV5VnR_AA
40479   qmymSqVwHYRqdwfcBatzpQ
78134   p5rpYtxS5xPQjt3MXYVEwA
84520   mz9ltimeAIy2c2qf5ctljw
92918   KskYqH1Bi7Z_61pH6Om8pg
93528   A0X1baHPgw9IiBRivu0G9g

                                                categories       city  \
104700                                [Pizza, Restaurants]  Las Vegas
110934  [Breakfast & Brunch, Mexican, Restaurants, Ame...  Las Vegas
142630  [Italian, Wine Bars, Restaurants, Nightlife, B...  Las Vegas
14551   [Food, Coffee & Tea, Breakfast & Brunch, Cafes...  Las Vegas
32230                          [Steakhouses, Restaurants]  Las Vegas
40479   [American (New), Restaurants, Sandwiches, Bars...  Las Vegas
78134   [Vegetarian, Restaurants, Burgers, Vegan, Amer...  Las Vegas
84520             [Arts & Entertainment, Performing Arts]  Las Vegas
92918   [Automotive, Car Dealers, Restaurants, Thai, N...  Las Vegas
93528               [Bakeries, French, Restaurants, Food]  Las Vegas

                                    name  review_count  stars state
104700                 Grimaldi's Pizzeria           431    4.0    NV
110934                Nacho Daddy Downtown           723    4.0    NV
142630                            Due Forni           446    4.0    NV
14551              Sambalatte Torrefazione           752    4.0    NV
32230                          Echo & Rig          1665    4.5    NV
40479                        Vintner Grill           571    4.0    NV
78134                  Greens and Proteins           600    4.0    NV
84520   Cirque du Soleil - The Beatles LOVE          1766    4.5    NV
92918                        Lotus of Siam          3838    4.0    NV
93528                    Patisserie Manon           598    4.0    NV
```

We see that this user really likes American Restaurants, Pizza, etc. He probably lives in Las vegas

Predict using basic averages

```
In [86]: get_reccomendations("---1lKK3aKOuomHnwAkAow",5,avg_preds_df)

Out[86]:         predicted rating                       address  \
         135187          5.660683          4627 E Ivy St, Ste 1
```

```
100304          5.660683    2960 S Durango Dr, Ste 112
107956          5.660683  7910 S Rainbow Blvd, Ste 110
87518           5.660683           10520 S Eastern Ave
143283          5.660683              7608 W Cactus Rd

                  business_id  \
135187  lHOPh4DiYSqj9UJBXAq8hQ
100304  56_j_lcGj5X9SpM2KzLm4A
107956  Hp8k_RpSIWSeJguyaQpfIw
87518   Wcuo6YmYj3xhCso5sMQcOw
143283  ZKsVCA89iXMccf3fEhS3iw

                                              categories       city  \
135187  [Home Services, Local Services, Self Storage, ...       Mesa
100304  [Laser Hair Removal, Beauty & Spas, Skin Care,...  Las Vegas
107956  [Gelato, Food, Desserts, Ice Cream & Frozen Yo...  Las Vegas
87518   [Pizza, Gluten-Free, Restaurants, Fast Food, S...  Henderson
143283  [Restaurants, Seafood, Cajun/Creole, American ...     Peoria

                                    name  review_count  stars state
135187  Just-In Time Moving and Delivery           374    5.0    AZ
100304       Fabulous Eyebrow Threading           453    5.0    NV
107956                       Gelatology           473    5.0    NV
87518           Blaze Fast-Fire'd Pizza           364    4.5    NV
143283                Angry Crab Peoria           365    4.5    AZ
```

Our baseline model recomends only one restaurant with American food and a few places with weird categories.

Predict using lasso regression

```
In [88]: get_reccomendations("---1lKK3aKOuomHnwAkAow",5,preds_array_reg_df)
```

```
Out[88]:        predicted rating               address             business_id  \
        36525          4.719550  3799 Las Vegas Blvd S  XnJeadLrlj9AZB8qSdIR2Q
        89310          4.710545         115 Federal St  X-b4-QvZLENnf3yFwhpSXQ
        15676          4.675836            Flamingo Rd  ty5KQYqYRxwXDG_e4pz-4w
        95839          4.672092  3600 S Las Vegas Blvd  NCFwm2-TDb-oBQ2medmYDg
        87314          4.649546                         jeTvVMOR8W_04xFsPjzOEQ

                                                categories       city  \
        36525                                [Restaurants, French]  Las Vegas
        89310  [Baseball Fields, Stadiums & Arenas, Active Li...  Pittsburgh
        15676            [Arts & Entertainment, Performing Arts]  Las Vegas
        95839  [Street Art, Performing Arts, Public Services ...  Las Vegas
        87314  [Local Services, Movers, Home Services, Self S...     Phoenix

                              name  review_count  stars state
```

```
36525         Joël  Robuchon           831    4.5    NV
89310                PNC Park           426    4.5    PA
15676                Absinthe          1452    4.5    NV
95839  Fountains of Bellagio           1083    4.5    NV
87314       Camelback Moving            394    5.0    AZ
```

Our baseline model based on lasso predicts restaurants completely unrelated to users preferences. Bad... At least it's mostly in Las Vegas

Predict using matrix factorization

```
In [136]: get_reccomendations("---1lKK3aKOuomHnwAkAow",5,fac_preds_df)
```

```
Out[136]:         predicted rating                          address  \
          151513          0.046371  5040 W Spring Mountain Rd, Ste 3
          98408           0.030232      2605 S Decatur Blvd, Ste 109
          131554          0.024574             3799 Las Vegas Blvd S
          129874          0.022182       1702 W Camelback Rd, Ste 14
          65918           0.020995           3989 Spring Mountain Rd

                             business_id  \
          151513  umXvdus9LbC6oxtLdXelFQ
          98408   OJdufUU3hVabgviIBHksYw
          131554  N0apJkxIem2E8irTBRKnHw
          129874  gQMAcDm8kv8ev7x2BshMwg
          65918   qkyCuFJF2Uboh6n2Lmuwlg

                                                    categories       city  \
          151513  [Wine Bars, Bars, Nightlife, Restaurants, Food...  Las Vegas
          98408   [Karaoke, Hawaiian, Restaurants, Bars, Nightlife]  Las Vegas
          131554  [Nightlife, Salad, Restaurants, Pizza, America...  Las Vegas
          129874                  [Vietnamese, Restaurants, Soup]    Phoenix
          65918          [Bakeries, Food, Sandwiches, Restaurants]  Las Vegas

                                               name  review_count  stars state
          151513                       Sweets Raku           749    4.5    NV
          98408                 Aloha Kitchen & Bar           425    3.5    NV
          131554  Wolfgang Puck Bar & Grill Las Vegas          1388    4.0    NV
          129874                          Pho Thanh           623    4.0    AZ
          65918                  Lee's Sandwiches            590    3.0    NV
```

Our Matrix Factorization model predicts places in Las vegas related to no nightlife, american food and surprisingly, vietnamese food.