

Team ID: PNT2022TMID23576

PROJECT NAME: DemandEst - AI powered Food Demand Forecaster

Team Leader

The image shows three vertically stacked screenshots of a Google Colab notebook titled "Data_preprocessing_and_Model_building (1).ipynb".

Code Cell 1:

```
plt.style.use('fivethirtyeight')
plt.figure(figsize=(12,7))
sns.distplot(trainfinal.num_orders, bins=25)
plt.xlabel("num_orders")
plt.ylabel("Number of Buyers")
plt.title("num_orders Distribution")
```

Output 1:

A histogram titled "num_orders Distribution" showing the distribution of the number of buyers. The x-axis ranges from 0 to 25,000 with major ticks every 5,000 units. The y-axis ranges from 0.0000 to 0.0040 with major ticks every 0.0005 units. The distribution is highly right-skewed, with the vast majority of buyers having 0 orders (peak at 0).

Code Cell 2:

```
trainfinal2 = trainfinal.drop(['id'], axis=1)
correlation = trainfinal2.corr(method='pearson')
columns = correlation.nlargest(8, 'num_orders').index
columns
```

Output 2:

```
Index(['num_orders', 'homepage_featured', 'emailer_for_promotion', 'op_area',
       'cuisine', 'city_code', 'region_code', 'category'],
      dtype='object')
```

```
correlation_map = np.corrcoef(trainfinal2[columns].values.T)
sns.set(font_scale=1.0)
heatmap = sns.heatmap(correlation_map, cbar=True, annot=True, square=True, fmt='.2f',
                      ticklabels=columns.values, xticklabels=columns.values)
plt.show()
```

Output 3:

A heatmap titled "correlation_map" showing the Pearson correlation coefficient between various features. The features on both axes are: num_orders, homepage_featured, emailer_for_promotion, op_area, cuisine, city_code, region_code, and category. The color scale ranges from -1.0 (dark purple) to 1.0 (dark orange). The diagonal elements are all 1.0. The highest correlations are seen between num_orders and other numerical features like homepage_featured (~0.29), emailer_for_promotion (~0.28), and op_area (~0.18).

Team Member 1

colab.research.google.com/drive/1mhHECd4iq-1OdkchHMYidY_hDhPD0lx.#scrollTo=MRaIP_RNUy5t

Data_preprocessing_and_Model_building (1).ipynb

```
File Edit View Insert Runtime Tools Help Last saved at November 19
```

Comment Share Connect Editing

+ Code + Text

plt.style.use('fivethirtyeight')
plt.figure(figsize=(12,7))
sns.distplot(trainfinal.num_orders, bins=25)
plt.xlabel("num_orders")
plt.ylabel("Number of Buyers")
plt.title("num_orders Distribution")

colab.research.google.com/drive/1mhHECd4iq-1OdkchHMYidY_hDhPD0lx.#scrollTo=MRaIP_RNUy5t

Data_preprocessing_and_Model_building (1).ipynb

```
File Edit View Insert Runtime Tools Help Last saved at November 19
```

Comment Share Connect Editing

+ Code + Text

/usr/local/lib/python3.7/dist-packages/seaborn/distributions.py:2619: FutureWarning: 'distplot' is a deprecated function and will be removed in a future version. Please adapt your code
warnings.warn(msg, FutureWarning)
Text(0.5, 1.0, 'num_orders Distribution')

num_orders Distribution

Number of Buyers

0.0040
0.0035
0.0030
0.0025
0.0020
0.0015
0.0010
0.0005
0.0000

0 5000 10000 15000 20000 25000

num_orders

colab.research.google.com/drive/1mhHECd4iq-1OdkchHMYidY_hDhPD0lx.#scrollTo=MRaIP_RNUy5t

Data_preprocessing_and_Model_building (1).ipynb

```
File Edit View Insert Runtime Tools Help Last saved at November 19
```

Comment Share Connect Editing

+ Code + Text

[] trainfinal2 = trainfinal.drop(['id'], axis=1)
correlation = trainfinal2.corr(method='pearson')
columns = correlation.nlargest(8, 'num_orders').index
columns

Index(['num_orders', 'homepage_featured', 'emailer_for_promotion', 'op_area',
 'cuisine', 'city_code', 'region_code', 'category'],
 dtype='object')

correlation_map = np.corrcoef(trainfinal2[columns].values.T)
sns.set(font_scale=1.0)
heatmap = sns.heatmap(correlation_map, cbar=True, annot=True, square=True, fmt='.2f',
 yticklabels=columns.values, xticklabels=columns.values)
plt.show()

num_orders 1.00 0.29 0.18 0.13 0.04 0.03 0.03
homepage_featured 0.29 1.00 0.39 0.04 -0.01 0.01 0.00 0.00
emailer_for_promotion 0.18 0.04 -0.05 1.00 0.01 0.13 0.02 0.01
op_area 0.13 -0.01 0.15 0.01 1.00 0.01 0.02 0.13
cuisine 0.04 0.01 -0.01 0.19 0.01 1.00 0.04 0.01
city_code 0.03 0.00 -0.01 0.02 0.02 0.04 1.00 0.01
region_code 0.03 0.00 0.10 0.01 0.13 0.01 0.01 1.00
category 0.03 0.00 0.10 0.01 0.13 0.01 0.01 1.00

num_orders homepage_featured emailer_for_promotion op_area cuisine city_code region_code category

Team Member 2

colab.research.google.com/drive/1mhHECd4iq-1OdkchHMYidY_hDhPD0lx.#scrollTo=MRaIP_RNUy5t

Data_preprocessing_and_Model_building (1).ipynb

```
File Edit View Insert Runtime Tools Help Last saved at November 19
```

Comment Share Connect Editing

+ Code + Text

plt.style.use('fivethirtyeight')
plt.figure(figsize=(12,7))
sns.distplot(trainfinal.num_orders, bins=25)
plt.xlabel("num_orders")
plt.ylabel("Number of Buyers")
plt.title("num_orders Distribution")

colab.research.google.com/drive/1mhHECd4iq-1OdkchHMYidY_hDhPD0lx.#scrollTo=MRaIP_RNUy5t

Data_preprocessing_and_Model_building (1).ipynb

```
File Edit View Insert Runtime Tools Help Last saved at November 19
```

Comment Share Connect Editing

+ Code + Text

/usr/local/lib/python3.7/dist-packages/seaborn/distributions.py:2619: FutureWarning: 'distplot' is a deprecated function and will be removed in a future version. Please adapt your code
warnings.warn(msg, FutureWarning)
Text(0.5, 1.0, 'num_orders Distribution')

num_orders Distribution

Number of Buyers

0.0040
0.0035
0.0030
0.0025
0.0020
0.0015
0.0010
0.0005
0.0000

0 5000 10000 15000 20000 25000

num_orders

colab.research.google.com/drive/1mhHECd4iq-1OdkchHMYidY_hDhPD0lx.#scrollTo=MRaIP_RNUy5t

Data_preprocessing_and_Model_building (1).ipynb

```
File Edit View Insert Runtime Tools Help Last saved at November 19
```

Comment Share Connect Editing

+ Code + Text

```
[ ] trainfinal2 = trainfinal.drop(['id'], axis=1)
correlation = trainfinal2.corr(method='pearson')
columns = correlation.nlargest(8, 'num_orders').index
columns

Index(['num_orders', 'homepage_featured', 'emailer_for_promotion', 'op_area',
       'cuisine', 'city_code', 'region_code', 'category'],
      dtype='object')
```

correlation_map = np.corrcoef(trainfinal2[columns].values.T)
sns.set(font_scale=1.0)
heatmap = sns.heatmap(correlation_map, cbar=True, annot=True, square=True, fmt='.2f',
 yticklabels=columns.values, xticklabels=columns.values)
plt.show()

num_orders 1.00 0.29 0.18 0.13 0.04 0.03 0.03
homepage_featured 0.29 1.00 0.39 0.04 -0.01 0.01 0.00 0.00
emailer_for_promotion 0.18 0.04 -0.05 1.00 0.01 0.13 0.02 0.01
op_area 0.13 -0.01 0.15 0.01 1.00 0.01 0.02 0.13
cuisine 0.04 0.01 -0.01 0.19 0.01 1.00 0.04 0.01
city_code 0.03 0.00 -0.01 0.02 0.02 0.04 1.00 0.01
region_code 0.03 0.00 0.10 0.01 0.13 0.01 0.01 1.00
category 0.03 0.00 0.10 0.01 0.13 0.01 0.01 1.00

num_orders homepage_featured emailer_for_promotion op_area cuisine city_code region_code category

Team Member 3

```
plt.style.use('fivethirtyeight')
plt.figure(figsize=(12,7))
sns.distplot(trainfinal.num_orders, bins=25)
plt.xlabel("num_orders")
plt.ylabel("Number of Buyers")
plt.title("num_orders Distribution")
```

```
[ ] trainfinal2 = trainfinal.drop(['id'], axis=1)
correlation = trainfinal2.corr(method='pearson')
columns = correlation.nlargest(8, 'num_orders').index
columns

Index(['num_orders', 'homepage_featured', 'emailer_for_promotion', 'op_area',
       'cuisine', 'city_code', 'region_code', 'category'],
      dtype='object')
```

```
correlation_map = np.corrcoef(trainfinal2[columns].values.T)
sns.set(font_scale=1.0)
heatmap = sns.heatmap(correlation_map, cbar=True, annot=True, square=True, fmt='.2f',
                      ticklabels=columns.values, xticklabels=columns.values)
plt.show()
```