```
['id', 'week', 'checkout_price', 'base_price', 'emailer_for_promotion', 'ho
mepage_featured', 'num_orders', 'category', 'cuisine', 'city_code', 'region
_code', 'center_type', 'op_area']
```

In [27]:

```
cols = cols[:2] + cols[9:] + cols[7:9] + cols[2:7]
print(cols)
```

```
['id', 'week', 'city_code', 'region_code', 'center_type', 'op_area', 'categ
ory', 'cuisine', 'checkout_price', 'base_price', 'emailer_for_promotion', '
homepage_featured', 'num_orders']
```

In [28]:

```
trainfinal = trainfinal[cols]
trainfinal.dtypes
```

Out[28]:

```
id                          int64
week                        int64
city_code                   int64
region_code                 int64
center_type                object
op_area                   float64
category                   object
cuisine                    object
checkout_price            float64
base_price                float64
emailer_for_promotion       int64
homepage_featured           int64
num_orders                  int64
dtype: object
```

In [30]:

```
from sklearn.preprocessing import LabelEncoder
```

In [32]:

```
lb1 = LabelEncoder()
trainfinal['center_type'] = lb1.fit_transform(trainfinal['center_type'])
lb2 = LabelEncoder()
trainfinal['category'] = lb1.fit_transform(trainfinal['category'])
lb3 = LabelEncoder()
trainfinal['cuisine'] = lb1.fit_transform(trainfinal['cuisine'])
```

In [33]:

```
trainfinal.head()
```

Out[33]:

TEAM ID: PNT2022TMID12176

	id	week	city_code	region_code	center_type	op_area	category	cuisine	checkout_price	base_price	emailer_for_promotion	homepage_featured	num_orders
0	1379560	1	647	56	2	2.0	0	3	136.83	152.29	0	0	177
1	1018704	2	647	56	2	2.0	0	3	135.83	152.29	0	0	323
2	1196273	3	647	56	2	2.0	0	3	132.92	133.92	0	0	96
3	1116527	4	647	56	2	2.0	0	3	135.86	134.86	0	0	163
4	1343872	5	647	56	2	2.0	0	3	146.50	147.50	0	0	215

```
trainfinal. shape
```

Out[34]:

```
(456548, 13)
```

In [36]:

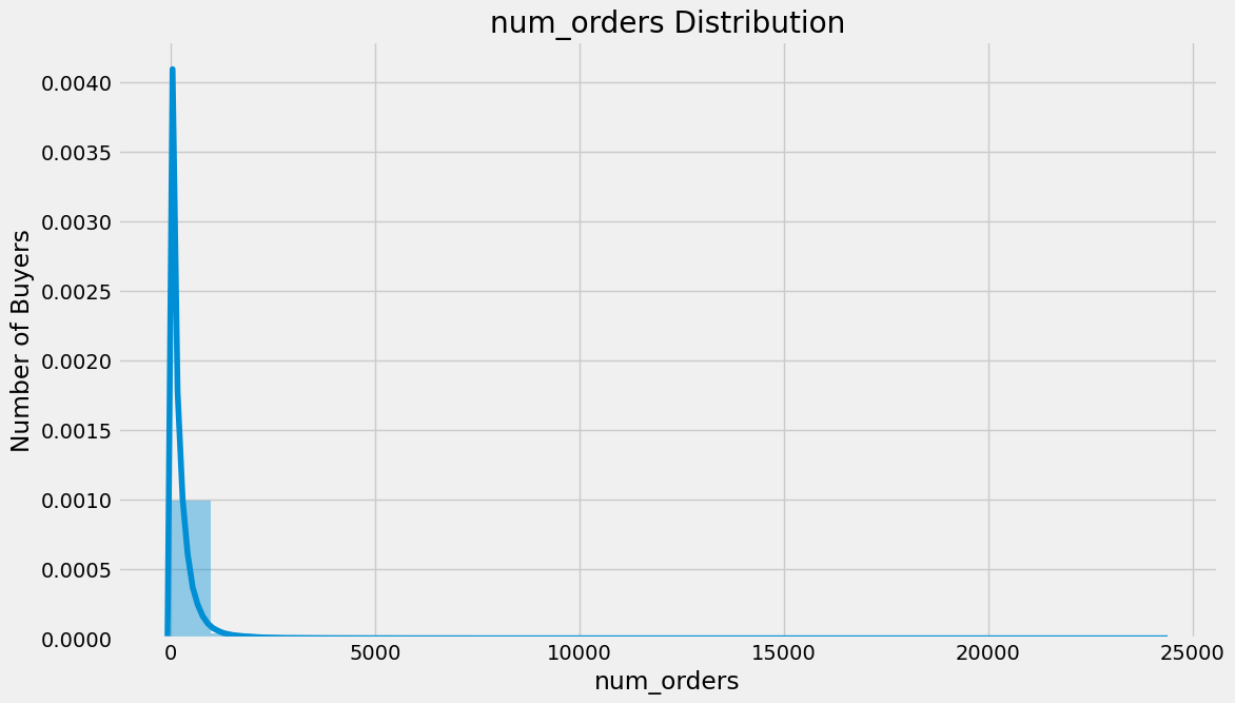
```
plt.style.use('fivethirtyeight')
plt.figure(figsize=(12,7))
sns.distplot(trainfinal.num_orders, bins = 25)
plt.xlabel("num_orders")
plt.ylabel("Number of Buyers")
plt.title("num_orders Distribution")
```

```
C:\Users\NANDHITHA\Anaconda3\lib\site-packages\seaborn\distributions.py:2619: FutureWarning: `distplot` is a deprecated function and will be removed in a future version. Please adapt your code to use either `displot` (a figure-level function with similar flexibility) or `histplot` (an axes-level function for histograms).
  warnings.warn(msg, FutureWarning)
```

Out[36]:

```
Text(0.5, 1.0, 'num_orders Distribution')
```

TEAM ID: PNT2022TMID12176



In [37]:

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

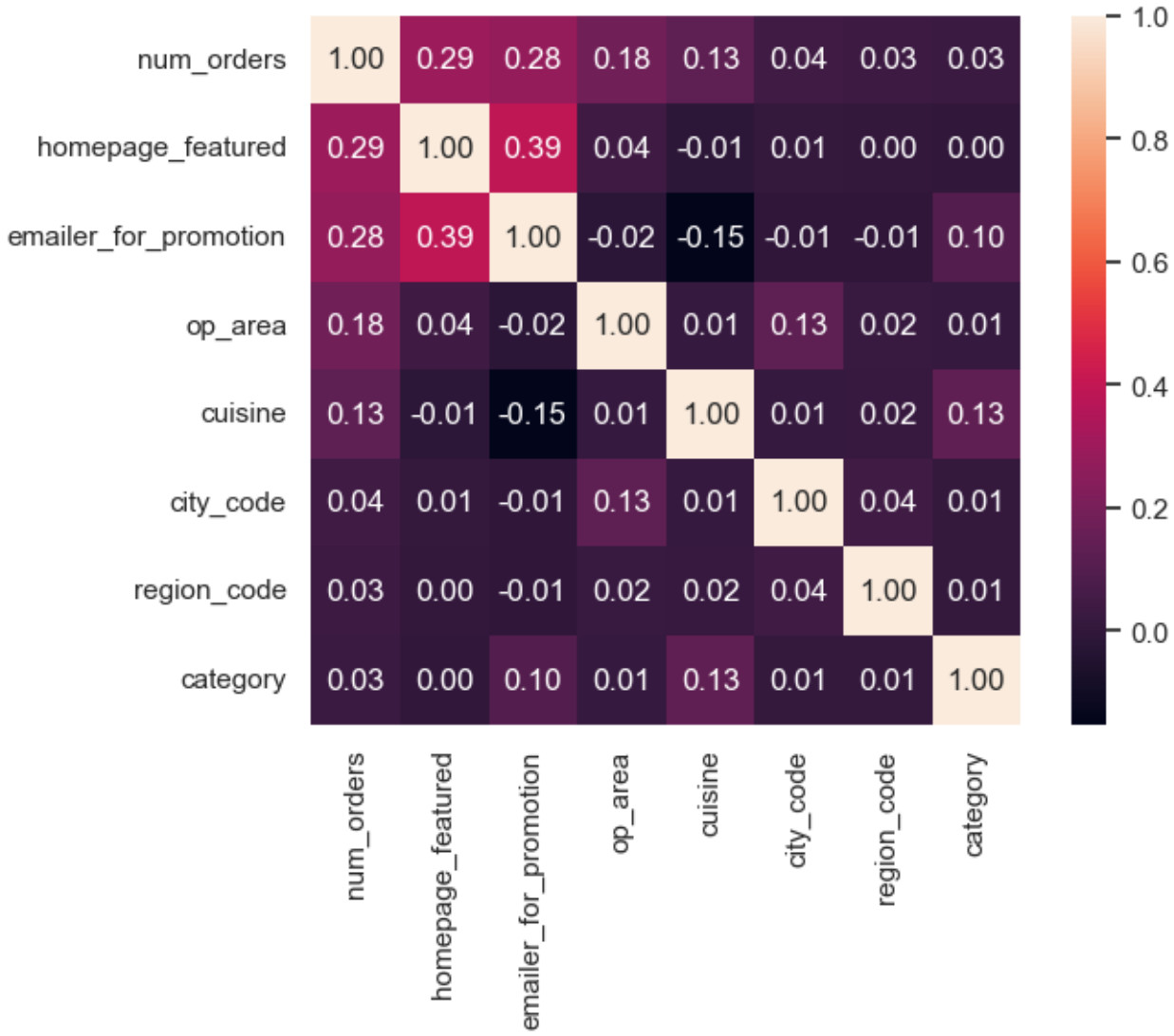
Out[37]:

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

In [39]:

```
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()
```

TEAM ID: PNT2022TMID12176



In [41]:

```
features = columns.drop(['num_orders'])
trainfinal3 = trainfinal[features]
X = trainfinal3.values
y = trainfinal[ 'num_orders'].values
```

In [42]:

```
trainfinal3.head()
```

Out[42]:

	homepage_featured	emailer_for_promotion	op_area	cuisine	city_code	region_code	category
0	0	0	2.0	3	647	56	0
1	0	0	2.0	3	647	56	0

TEAM ID: PNT2022TMID12176

	homepage_featured	emailer_for_promotion	op_area	cuisine	city_code	region_code	category
2	0	0	2.0	3	647	56	0
3	0	0	2.0	3	647	56	0
4	0	0	2.0	3	647	56	0

In [45]:

```
from sklearn.model_selection import train_test_split
```

In [47]:

```
X_train, X_val, y_train, y_val = train_test_split(X, y, test_size=0.25)
```