Texell Case

February 17, 2020

[1]:
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns

#Ref https://datascience.stackexchange.com/questions/25596/
 ↪how-to-plot-two-columns-of-single-dataframe-on-y-axis
```

[2]:
```
data = pd.read_excel('Texxell Dataset.xlsx')

data['week'] = data['week'].apply(lambda x : x[5:])
data['week'] = data['week'].astype(int)
data.sample(5)
```

[2]:

	date	week	volume_pre	volume_reg	year	dow
60222	2010-09-26	39	8679.13	12558.120	2010	sunday
164491	2009-04-17	16	3986.84	5768.688	2009	friday
155076	2009-06-03	22	1168.34	1534.030	2009	wednesday
106788	2010-02-03	5	12762.84	18466.970	2010	wednesday
27195	2011-03-12	11	2926.85	4234.954	2011	saturday

	price_reg_texon	price_reg_other	price_reg_branded	price_pre_texon
60222	2.514	NaN	NaN	2.754
164491	1.984	1.994	NaN	2.274
155076	2.814	NaN	NaN	3.054
106788	2.514	2.524	2.479	2.834
27195	3.514	3.464	NaN	3.914

	…	m9	m10	m11	m12	weekend	Pdiff_pre_other	Pdiff_pre_branded
60222	…	1	0	0	0	0	NaN	NaN
164491	…	0	0	0	0	1	-0.025	NaN
155076	…	0	0	0	0	0	NaN	NaN
106788	…	0	0	0	0	0	0.085	0.16
27195	…	0	0	0	0	1	0.055	NaN

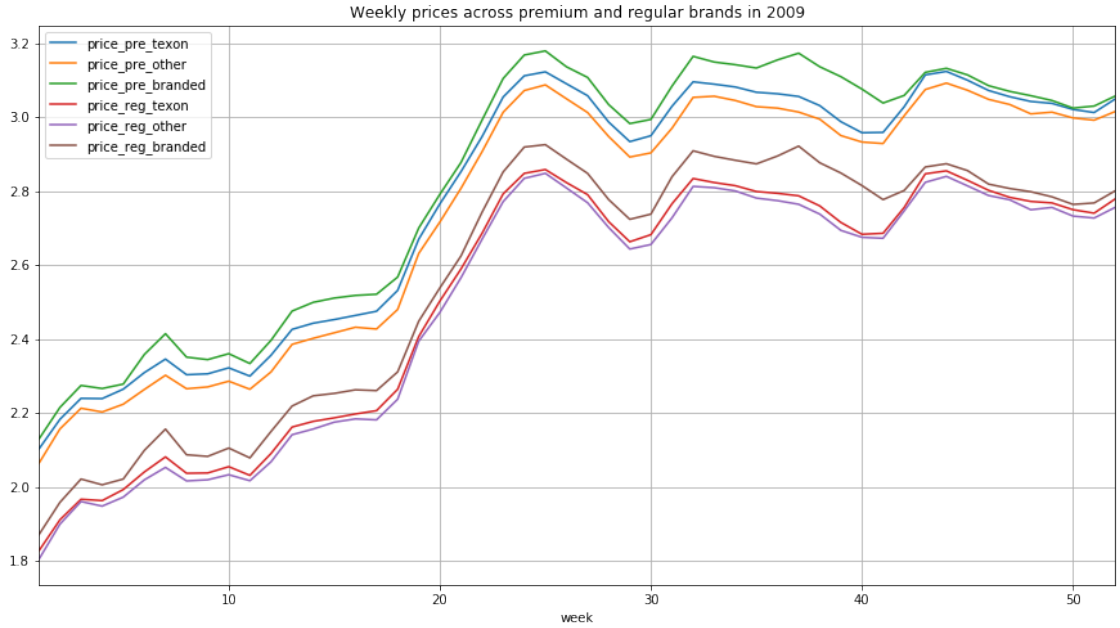
	Pdiff_reg_other	Pdiff_reg_branded	over4
60222	NaN	NaN	0

```
164491            -0.01               NaN      0
155076              NaN               NaN      0
106788            -0.01             0.035      0
27195              0.05               NaN      0

[5 rows x 40 columns]
```

[3]:
```
df = data.groupby(['year','week'])['price_pre_texon'].mean().
 ↪to_frame('price_pre_texon').reset_index()
temp = data.groupby(['year','week'])['price_pre_other'].mean().
 ↪to_frame('price_pre_other').reset_index()
df = df.merge(temp)
temp = data.groupby(['year','week'])['price_pre_branded'].mean().
 ↪to_frame('price_pre_branded').reset_index()
df = df.merge(temp)

temp = data.groupby(['year','week'])['price_reg_texon'].mean().
 ↪to_frame('price_reg_texon').reset_index()
df = df.merge(temp)
temp = data.groupby(['year','week'])['price_reg_other'].mean().
 ↪to_frame('price_reg_other').reset_index()
df = df.merge(temp)
temp = data.groupby(['year','week'])['price_reg_branded'].mean().
 ↪to_frame('price_reg_branded').reset_index()
df = df.merge(temp)
```

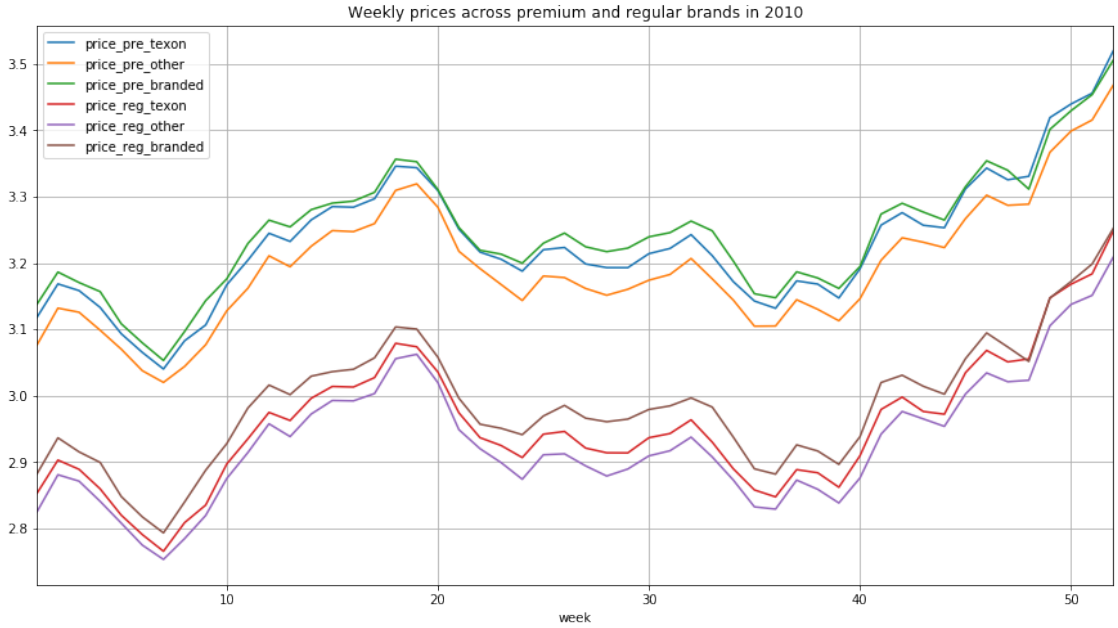
[4]:
```
df[df['year'] == 2009].plot(x='week', y=['price_pre_texon', 'price_pre_other',␣
 ↪'price_pre_branded',
                                         'price_reg_texon', 'price_reg_other',␣
 ↪'price_reg_branded'],
                            figsize=(15,8),
                            title = 'Weekly prices across premium and regular␣
 ↪brands in 2009',
                            grid=True)
```

[4]:
```
<matplotlib.axes._subplots.AxesSubplot at 0x1a1a0a5e90>
```



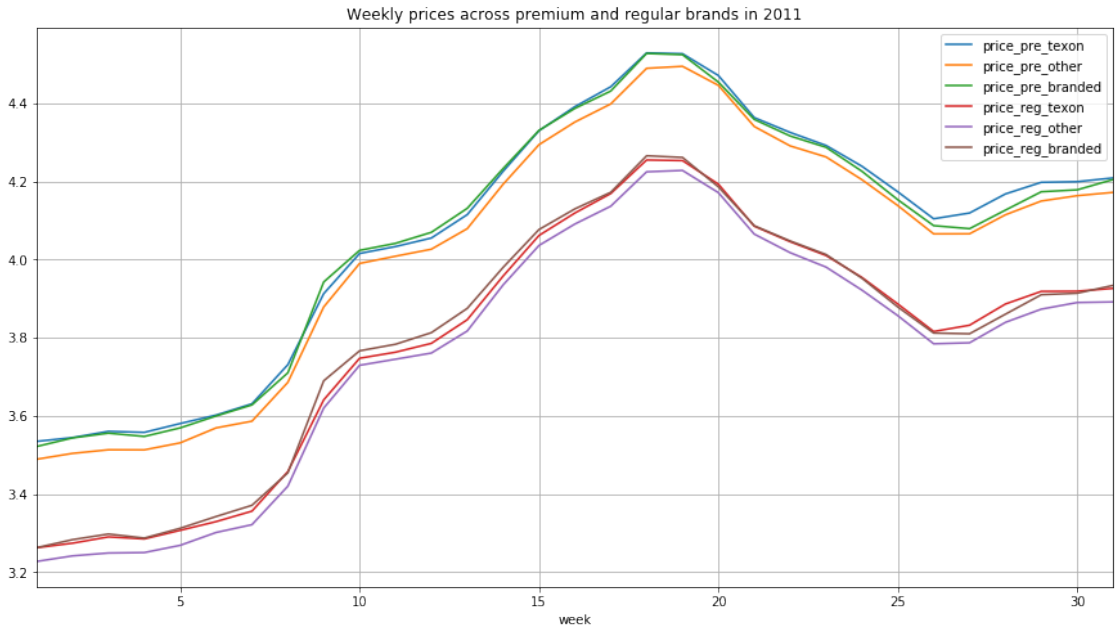
[5]:
```
df[df['year'] == 2010].plot(x='week', y=['price_pre_texon', 'price_pre_other',␣
↪'price_pre_branded',
                                         'price_reg_texon', 'price_reg_other',␣
↪'price_reg_branded'],
                            figsize=(15,8),
                             title = 'Weekly prices across premium and regular␣
↪brands in 2010',
                            grid=True)
```

[5]: <matplotlib.axes._subplots.AxesSubplot at 0x1a2e46f290>



[6]:
```
df[df['year'] == 2011].plot(x='week', y=['price_pre_texon', 'price_pre_other',␣
↪'price_pre_branded',
                                         'price_reg_texon', 'price_reg_other',␣
↪'price_reg_branded'],
                            figsize=(15,8),
                             title = 'Weekly prices across premium and regular␣
↪brands in 2011',
                            grid=True)
```

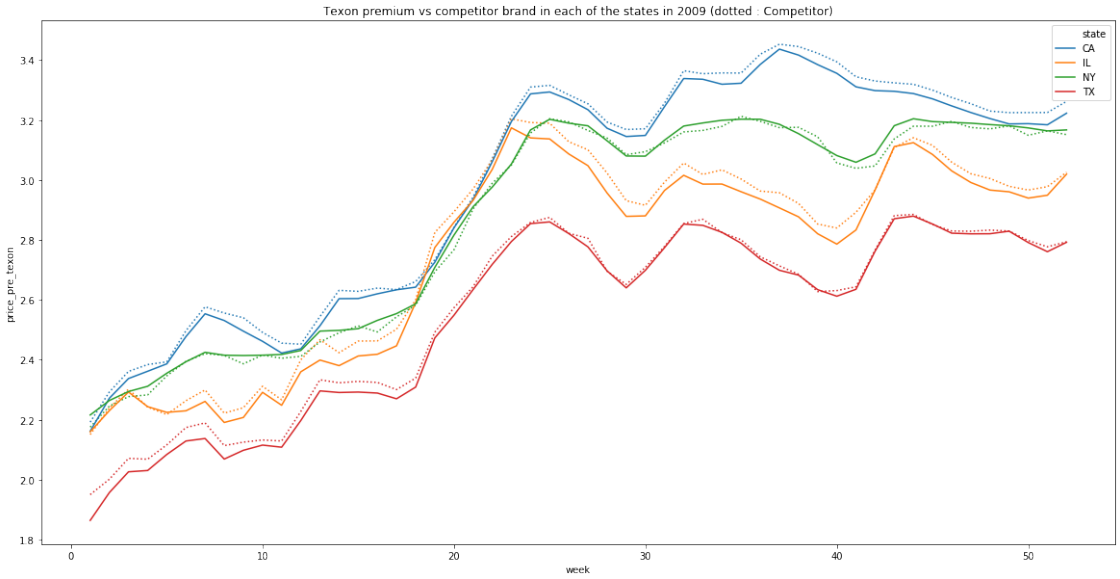
[6]: <matplotlib.axes._subplots.AxesSubplot at 0x1a2ab675d0>



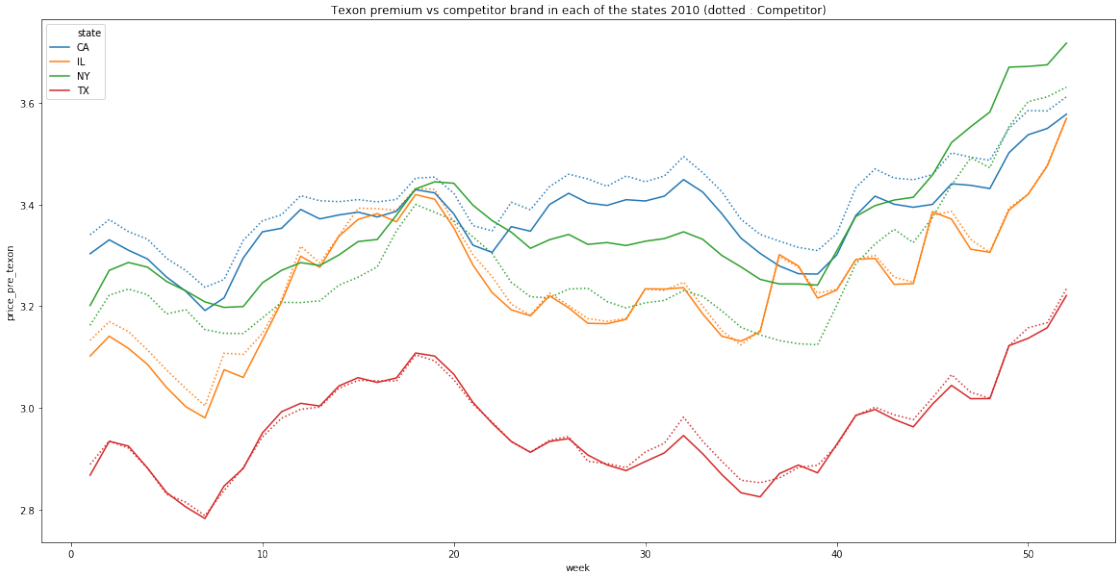
[8]:
```
df = data.groupby(['year','week','state'])['price_pre_texon'].mean().
 ↪to_frame('price_pre_texon').reset_index()
temp = data.groupby(['year','week','state'])['price_pre_other'].mean().
 ↪to_frame('price_pre_other').reset_index()
df = df.merge(temp)
temp = data.groupby(['year','week','state'])['price_pre_branded'].mean().
 ↪to_frame('price_pre_branded').reset_index()
df = df.merge(temp)

temp = data.groupby(['year','week','state'])['price_reg_texon'].mean().
 ↪to_frame('price_reg_texon').reset_index()
df = df.merge(temp)
temp = data.groupby(['year','week','state'])['price_reg_other'].mean().
 ↪to_frame('price_reg_other').reset_index()
df = df.merge(temp)
temp = data.groupby(['year','week','state'])['price_reg_branded'].mean().
 ↪to_frame('price_reg_branded').reset_index()
df = df.merge(temp)
```

[9]:
```
plt.figure(figsize=(20,10))

ax1 = sns.lineplot(data=df[df['year'] == 2009], x='week',␣
 ↪y='price_pre_branded', hue='state')

ax2 = sns.lineplot(data=df[df['year'] == 2009], x='week', y='price_pre_texon',␣
 ↪hue='state', legend=False)
```

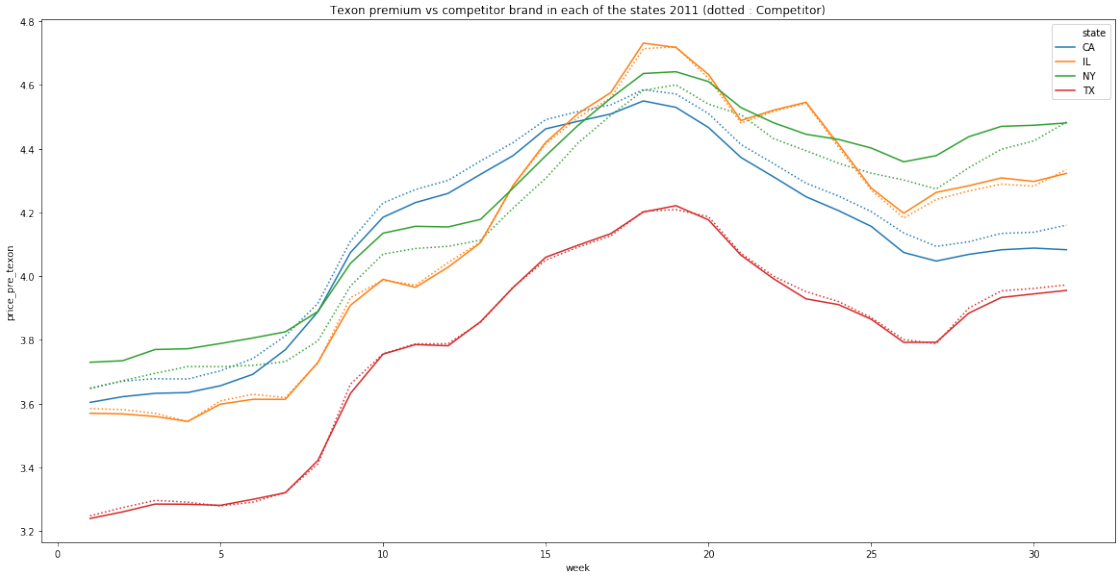
```
ax2.lines[0].set_linestyle('dotted')
ax2.lines[1].set_linestyle('dotted')
ax2.lines[2].set_linestyle('dotted')
ax2.lines[3].set_linestyle('dotted')

plt.title('Texon premium vs competitor brand in each of the states in 2009␣
 ↪(dotted : Competitor)')
plt.show()
```



[10]:
```
plt.figure(figsize=(20,10))

ax1 = sns.lineplot(data=df[df['year'] == 2010], x='week',␣
 ↪y='price_pre_branded', hue='state')

ax2 = sns.lineplot(data=df[df['year'] == 2010], x='week', y='price_pre_texon',␣
 ↪hue='state', legend=False)

ax2.lines[0].set_linestyle('dotted')
ax2.lines[1].set_linestyle('dotted')
ax2.lines[2].set_linestyle('dotted')
ax2.lines[3].set_linestyle('dotted')

plt.title('Texon premium vs competitor brand in each of the states 2010 (dotted␣
 ↪: Competitor)')
plt.show()
```

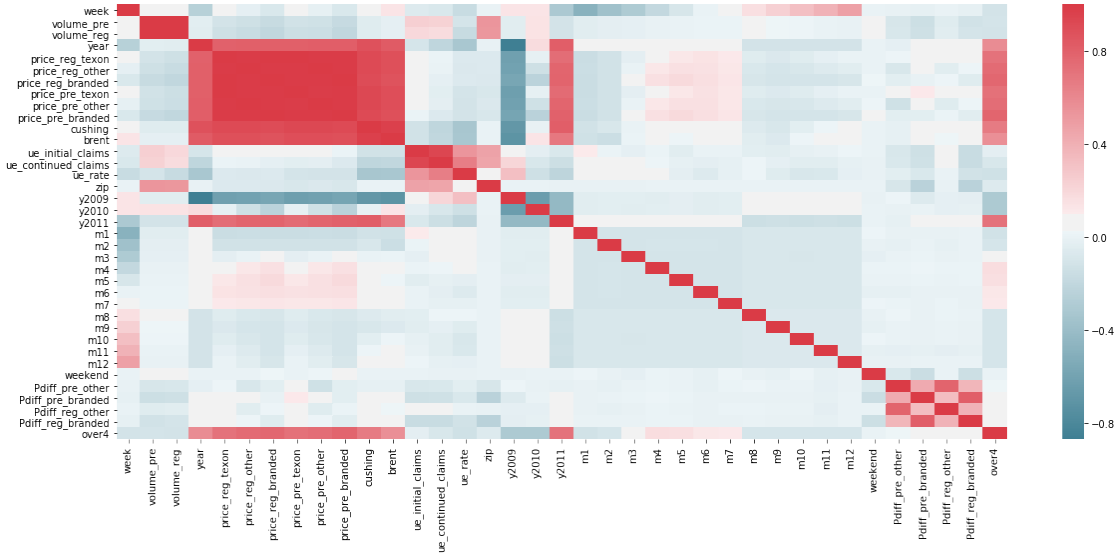


[11]:
```
plt.figure(figsize=(20,10))

ax1 = sns.lineplot(data=df[df['year'] == 2011], x='week',␣
↪y='price_pre_branded', hue='state')

ax2 = sns.lineplot(data=df[df['year'] == 2011], x='week', y='price_pre_texon',␣
↪hue='state', legend=False)

ax2.lines[0].set_linestyle('dotted')
ax2.lines[1].set_linestyle('dotted')
ax2.lines[2].set_linestyle('dotted')
ax2.lines[3].set_linestyle('dotted')

plt.title('Texon premium vs competitor brand in each of the states 2011 (dotted␣
↪: Competitor)')
plt.show()
```



[16]:
```
#Ref : https://stackoverflow.com/questions/29432629/
 ↪plot-correlation-matrix-using-pandas

f, ax = plt.subplots(figsize=(20, 8))
corr = data.corr()
sns.heatmap(corr,
            mask=np.zeros_like(corr, dtype=np.bool),
            cmap=sns.diverging_palette(220, 10, as_cmap=True),
            ax=ax)
```

[16]: <matplotlib.axes._subplots.AxesSubplot at 0x1a2de72b10>



[ ]:

[ ]: