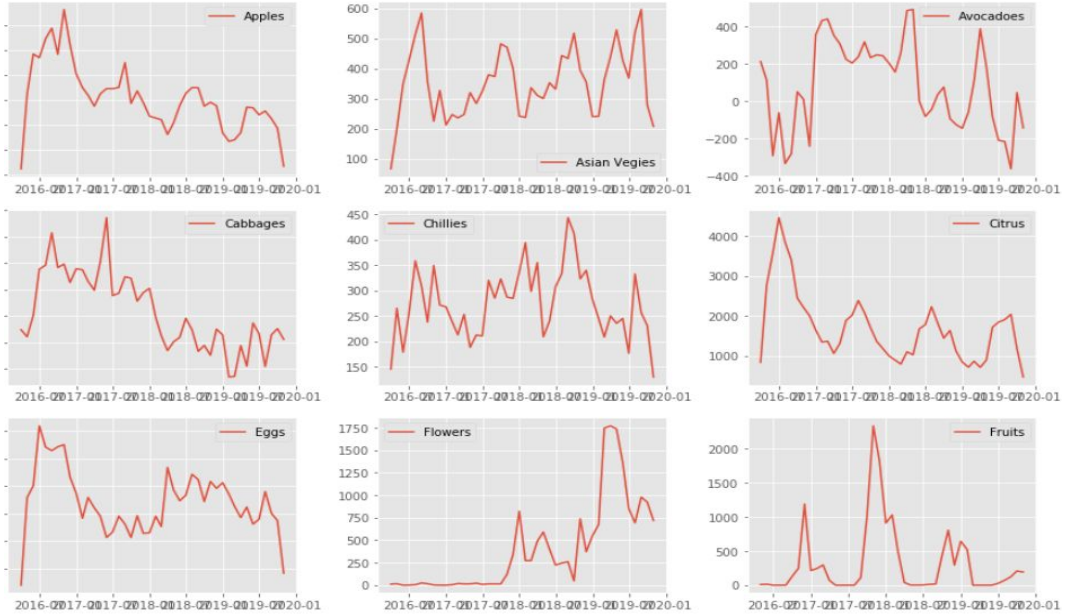
Grocery sales Data By Category

By Olga Illarionova

Profits by categories aggregated by month

For first look there are no clear trends, biu may be there ais seasonality in data



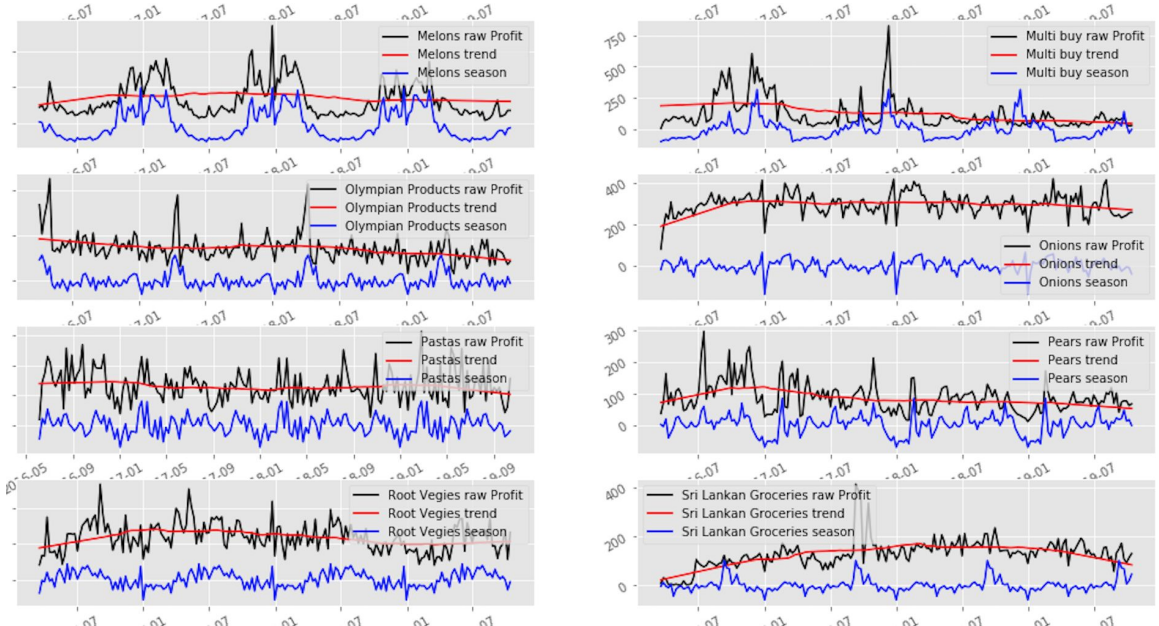
Highest/lowest selling products by category

To familiarize myself with data I calculated highest/lowest selling product by each category by month.

	DATE	CATEGORY	NAME	UNITS	TOTAL_PRICESELL	PROFIT
50	2016-04-30	Grapes	Grapes black	303.914	1212.61686	1212.61686
51	2016-04-30	Groceries-Dry Goods	Monsoon Basmathi 5kg	8.000	79.92000	79.92000
52	2016-04-30	Herbs	Dill	65.000	195.00000	195.00000
53	2016-04-30	Kalamata olives	Virgin Olive oil 4Lt	4.000	159.96000	21.96000
54	2016-04-30	Lettuces	Lettuce Iceberg	240.000	600.00000	600.00000
55	2016-04-30	Melons	Rockmelon	142.000	255.60000	255.60000
56	2016-04-30	Multi buy	kiwifruit multibuy 5 for $2	59.000	118.00000	94.99000
57	2016-04-30	Mushrooms	Mushroom Cup	53.722	482.96078	214.35078
58	2016-04-30	Nuts	Cannellini Beans 1Kg	14.000	76.86000	18.06000

Trends 2

I use seasonal_decompose from stats package to look on trends. So most categories go slightly down or stay same



Look thoroughly on ‘Pears’ category

Train several Arima models on subset with category Pears,

Tried aggregate data by months, and by weeks.

Tried to predict Profit, and tried to predict amount of sales and price.

Finished with weeks and profits

Tuneparameters for model

Among several Arima model I choose those, with best AIC score

```
parameters = [
    {'AR': 5, 'MA': 0},
    {'AR': 3, 'MA': 1},
    {'AR': 4, 'MA': 0},
    {'AR': 4, 'MA': 1},
    {'AR': 5, 'MA': 1}
]

models = {}
for params in parameters:
    models[(params['AR'], params['MA'])] = ARIMA(pears_prof_week, order=(params['AR'], 1, params['MA'])).fit()

for model_params in models:
    print("ARIMA({}, 1, {}) AIC: {}".format(model_params[0], model_params[1], models[model_params].aic))
```

```
ARIMA(5, 1, 0) AIC: 1930.8717278803456
ARIMA(3, 1, 1) AIC: 1916.705344514336
ARIMA(4, 1, 0) AIC: 1929.486037454024
ARIMA(4, 1, 1) AIC: 1917.899154604852
ARIMA(5, 1, 1) AIC: 1918.8169031065017
```

How model works with all categories? Train data.

Fit ARIMA model with same parameters for all categories

on the dataset without **last 10 weeks.**

To make sure it’s not completely bad, compare it with very simple prediction : take just last week profit.

I calculate Mean Squared Error for Arima Models, and for that simple prediction, and find that in most cases Arima works better

Finally predict future

	CATEGORY	PROFIT
0	Apples	16759.544629
1	Asian Vegies	5706.783730
2	Avocadoes	-1142.276737
3	Bananas	-74.412576
4	Berries	2640.813535

Whats next?

- Check those categories, where my model performs bad
- Make prediction for every single product