Time Series Prediction for Wal-Mart Sales

Augusto Perez, James Ghosn, German Baltazar

06/29/2022

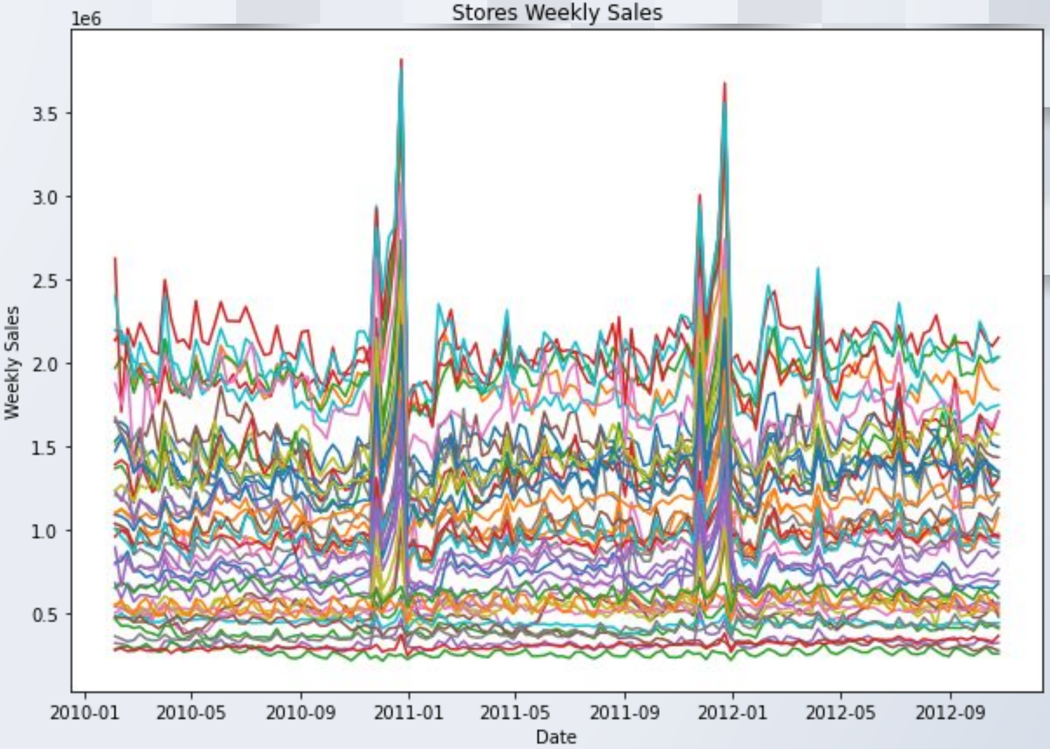
Data Description

- Weekly information regarding Wal-Mart sales
- Dataset with 421,570 observations
- Sum of sales per department

Date	Store	Dept	Weekly_Sales	Temperature
Fuel_Price	MarkDown1	MarkDown2	MarkDown3	MarkDown4
MarkDown5	CPI	Unemployment	IsHoliday	

EDA

- ❑ General observation shows little periodic behavior on the weekly sales
- ❑ Too many models to predict each store individually
- ❑ Use of K-Means to cluster the data into smaller groups

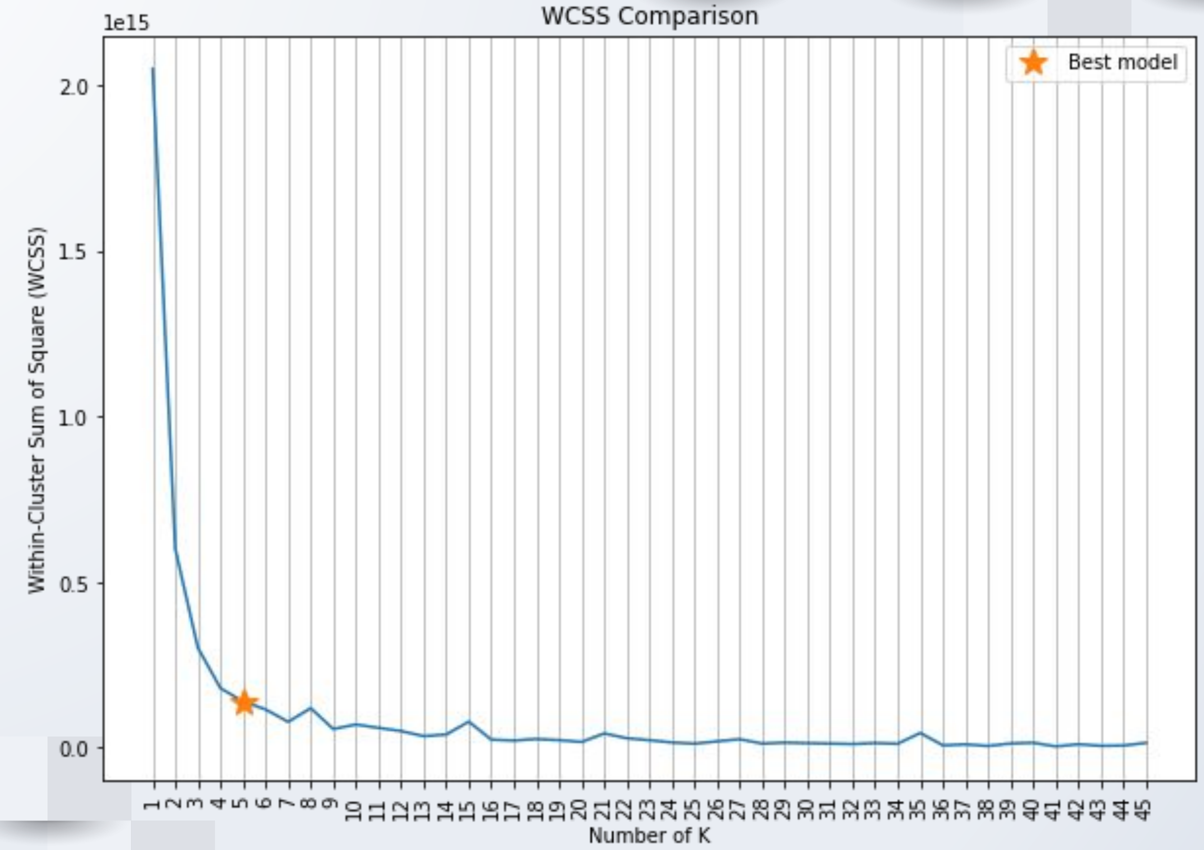


Elbow Method

- Vary the number of clusters k
- Obtain the WCSS (Within-Cluster Sum of Square)
- Stay with the model that breaks the drastic change on the WCSS (the elbow)

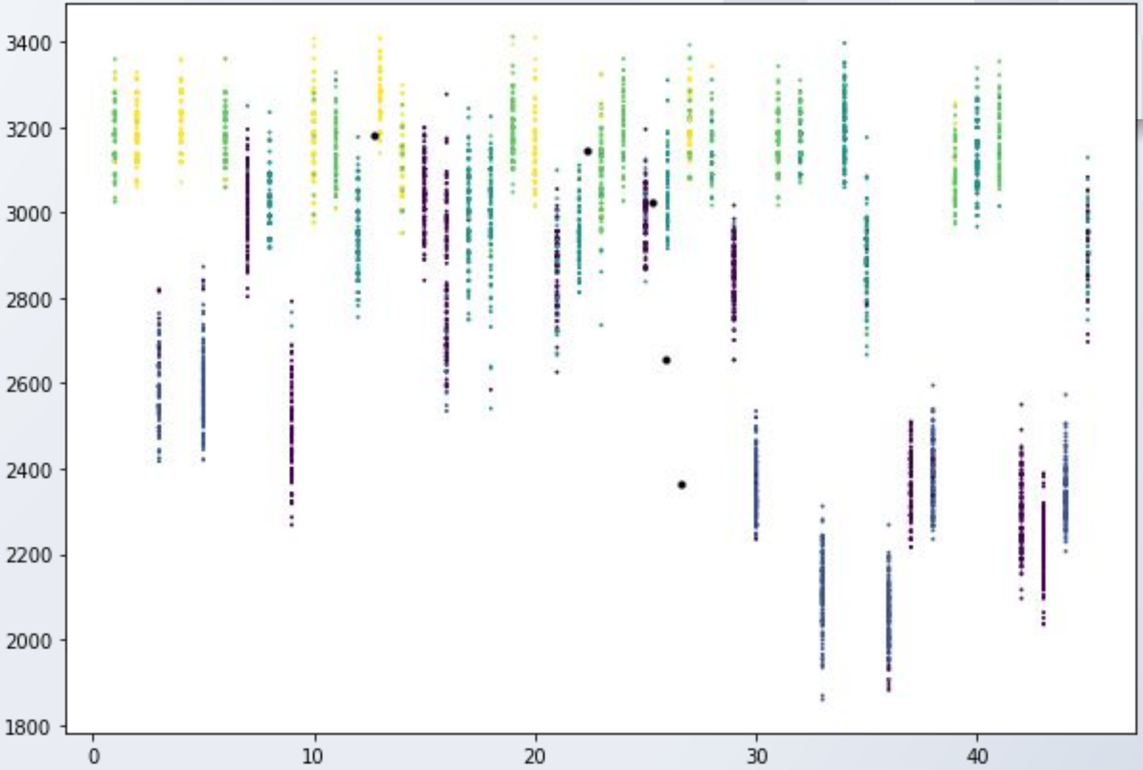
$$D_r = \sum_{i=1}^{n_r - 1} \sum_{j=i}^{n_r} ||d_i - d_j||_2$$

Elbow Method



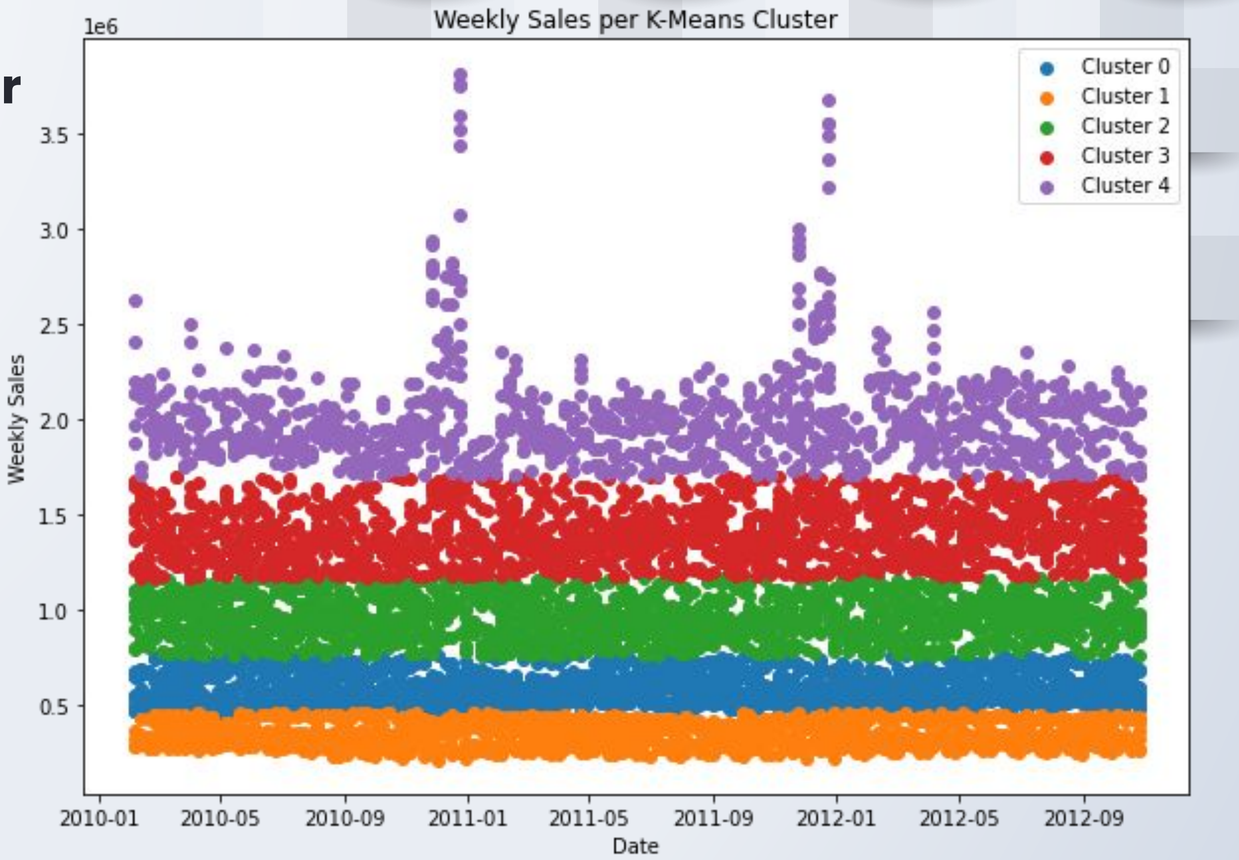
K-Means Visualization

- One possible way of observing
- Difficult to see multidimensional behavior on 2D



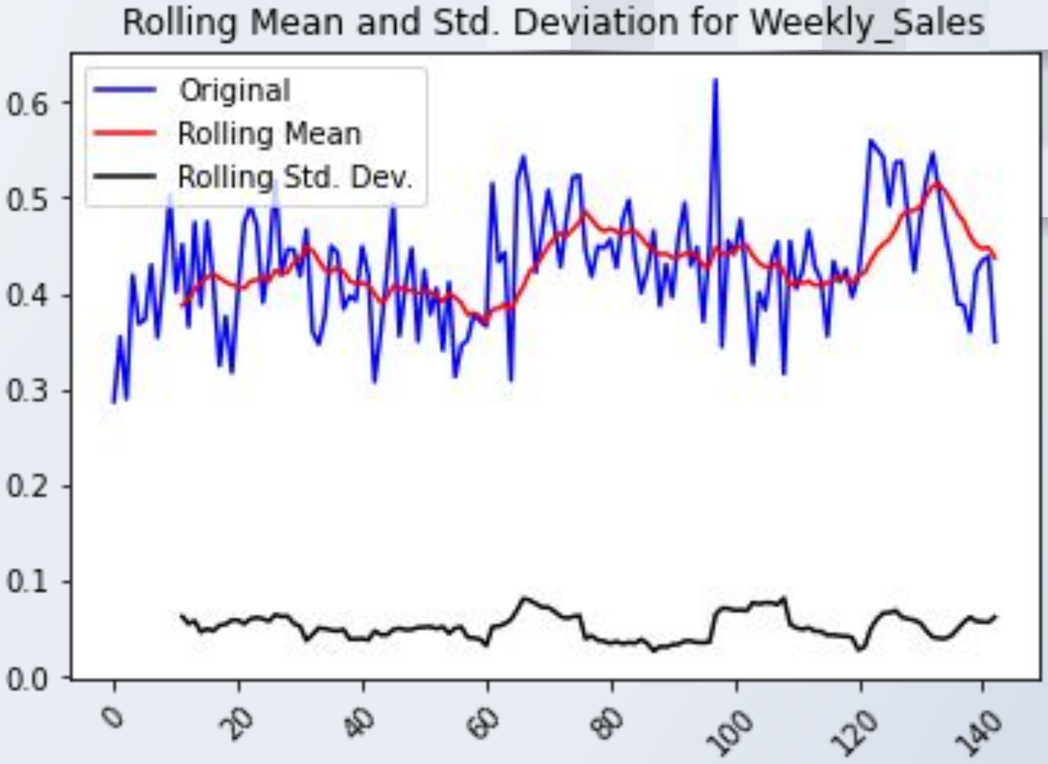
Weekly Sales per Cluster

- Classification based on range of prices

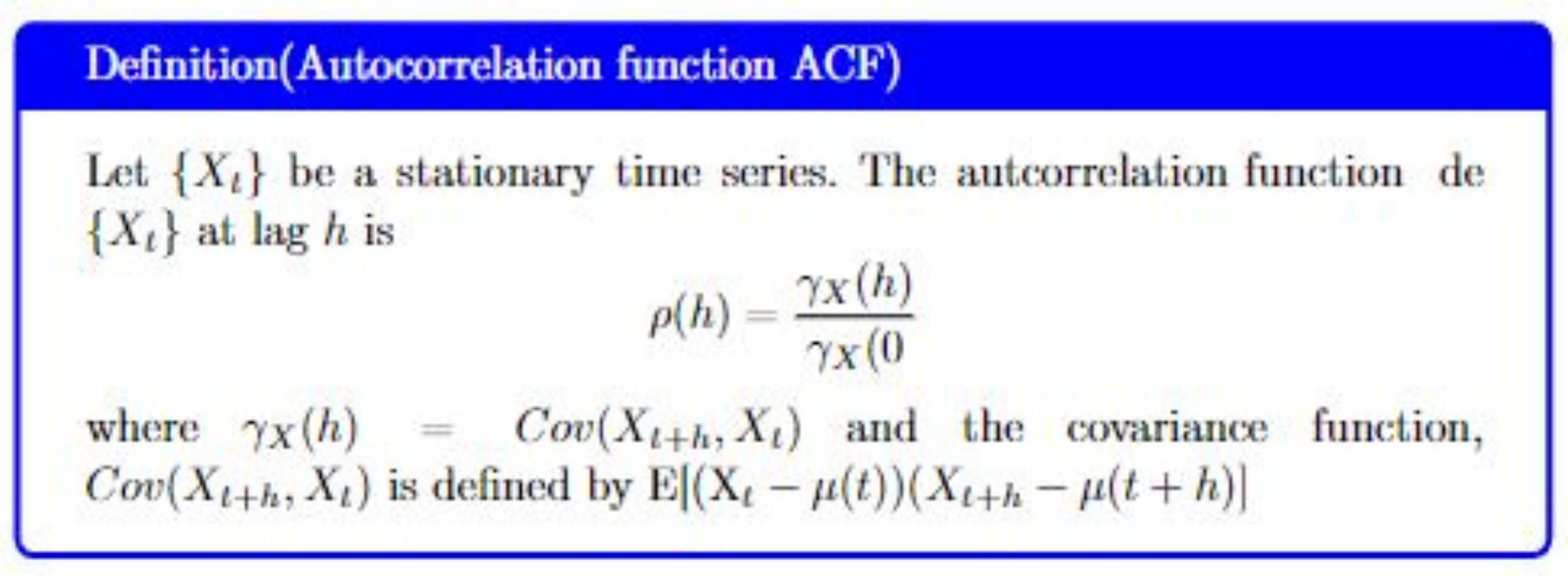


ARIMA

- Evaluate for stationarity of clusters
- Clusters 1 and 3 were non-stationary
- Applied differentiation for stationarity



ACF and PCF



Definition(Autocorrelation function ACF)

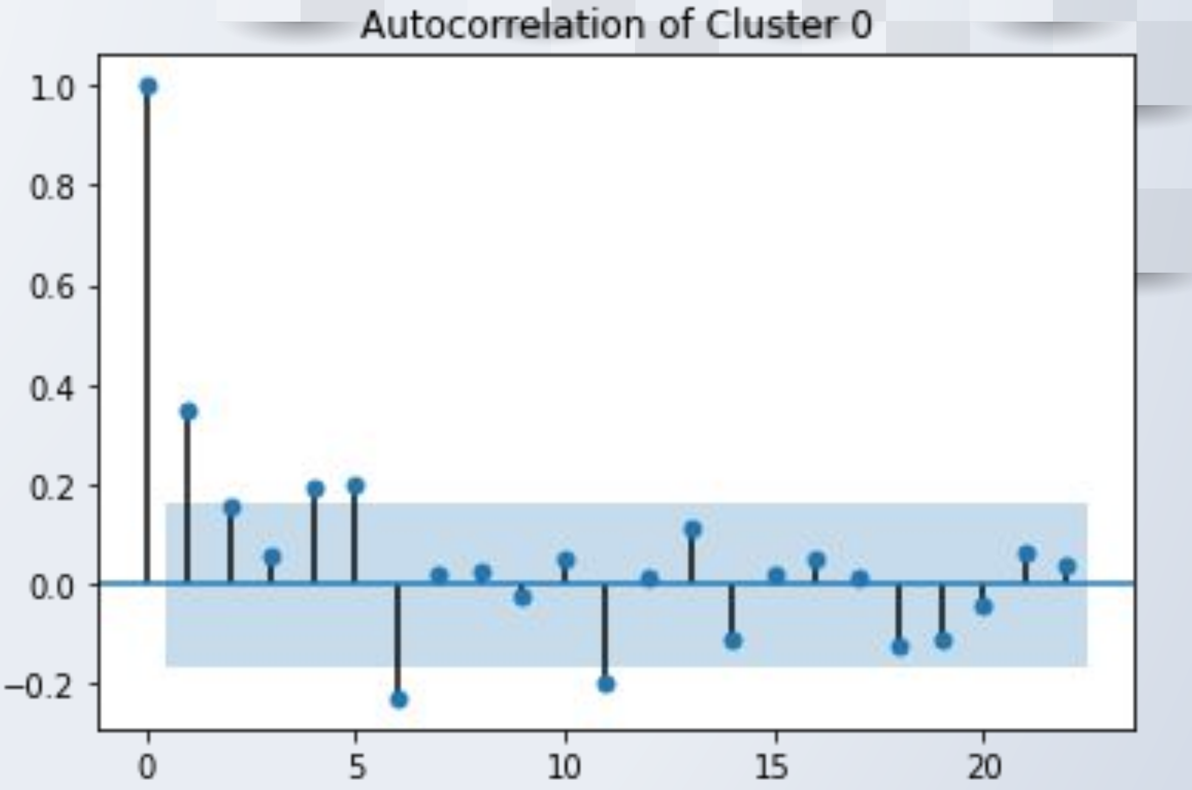
Let $\{X_t\}$ be a stationary time series. The autcorrelation function de $\{X_t\}$ at lag h is

$$\rho(h) = \frac{\gamma_X(h)}{\gamma_X(0}$$

where $\gamma_X(h) = Cov(X_{t+h}, X_t)$ and the covariance function, $Cov(X_{t+h}, X_t)$ is defined by $\mathrm{E}[(\mathrm{X}_t - \mu(t))(X_{t+h} - \mu(t+h)]$

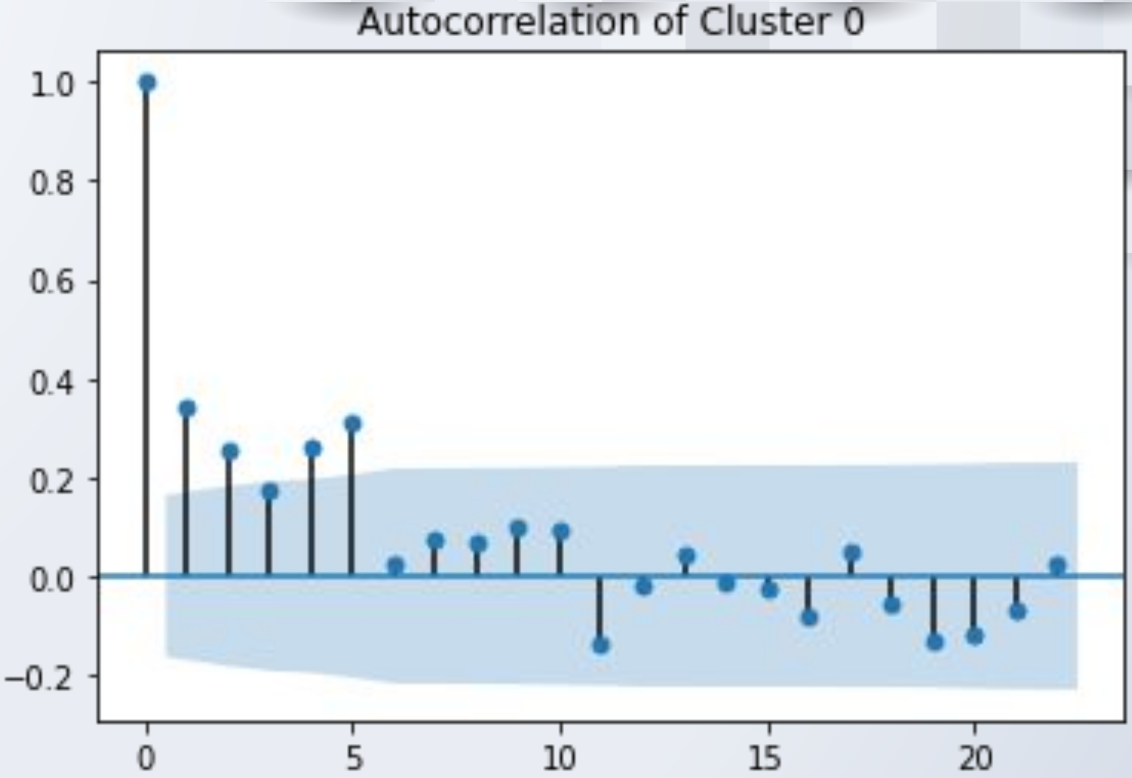
PACF for ARIMA

- ❑ Selected the smallest lag value closer to the decision boundary (without touching it)
- ❑ Determines the p parameter for ARIMA model



ACF for ARIMA

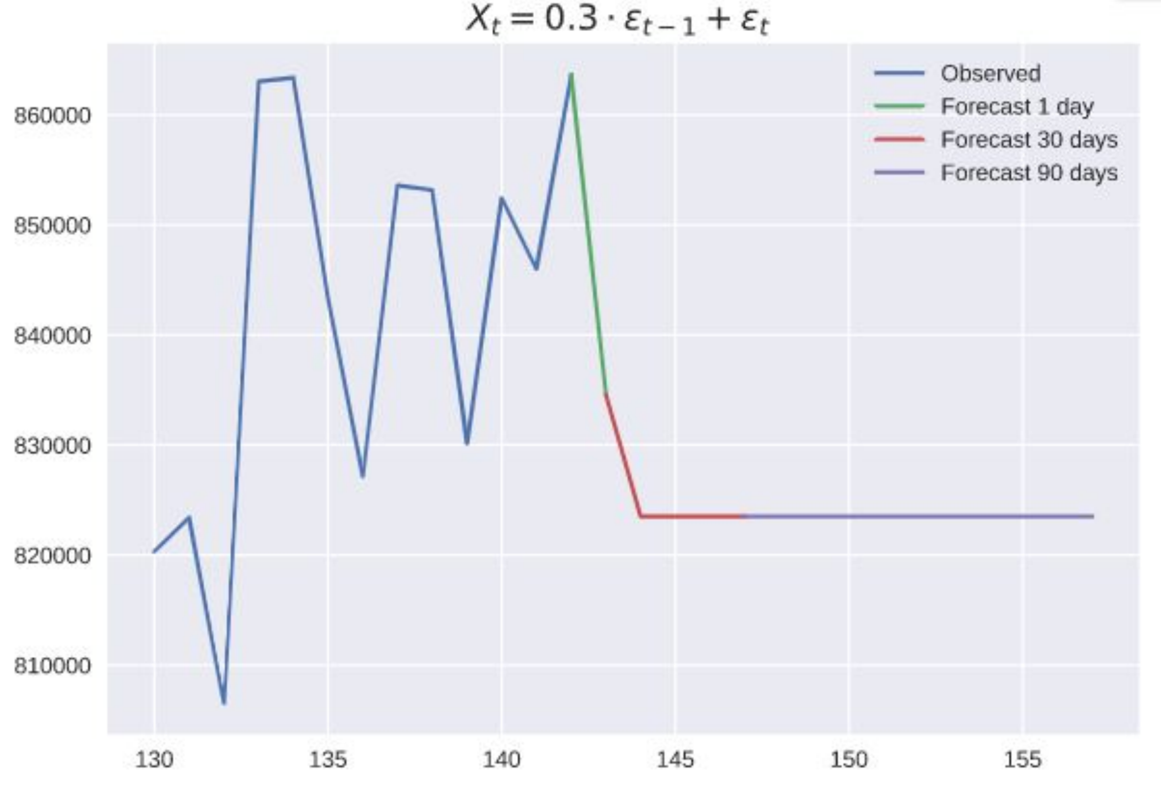
- Repeat same process for the q parameter



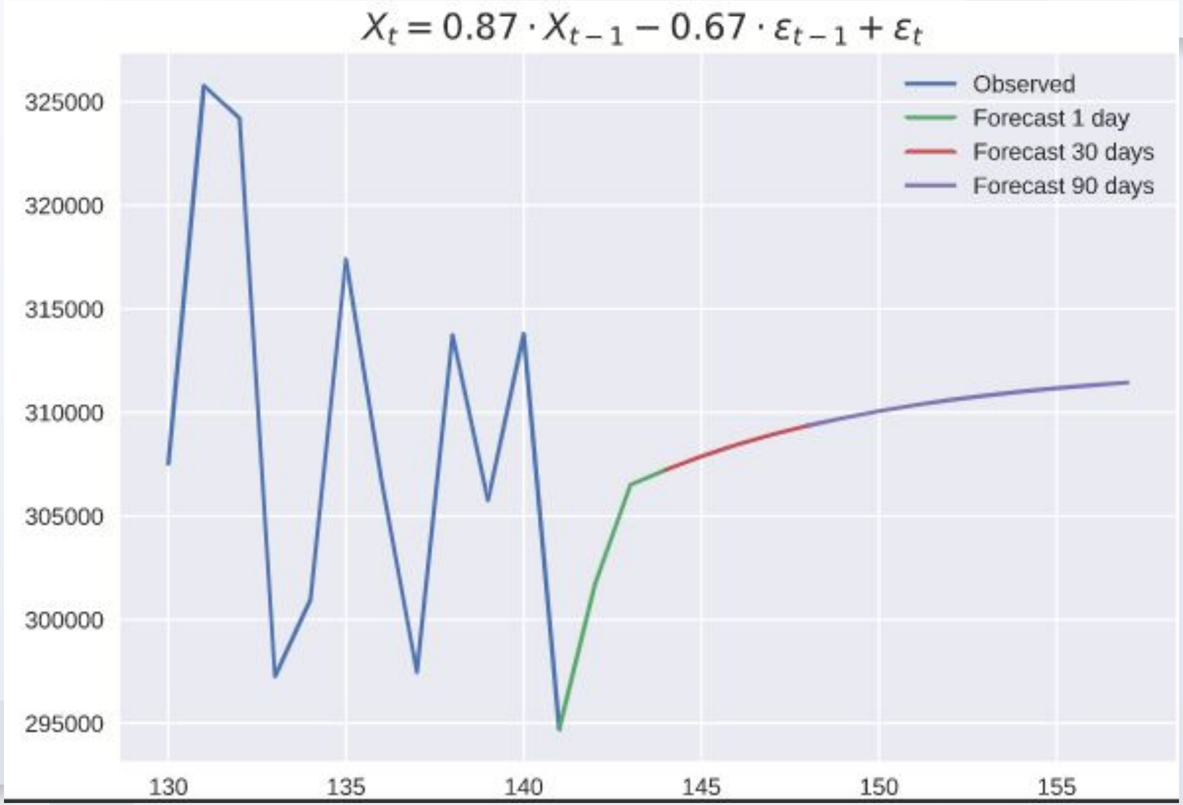
Selected ARIMA Structures

	p	d	q
Cluster 0	0	1	1
Cluster 1	3	0	4
Cluster 2	0	1	1
Cluster 3	1	0	1
Cluster 4	1	0	5

ARIMA Forecast Cluster 0

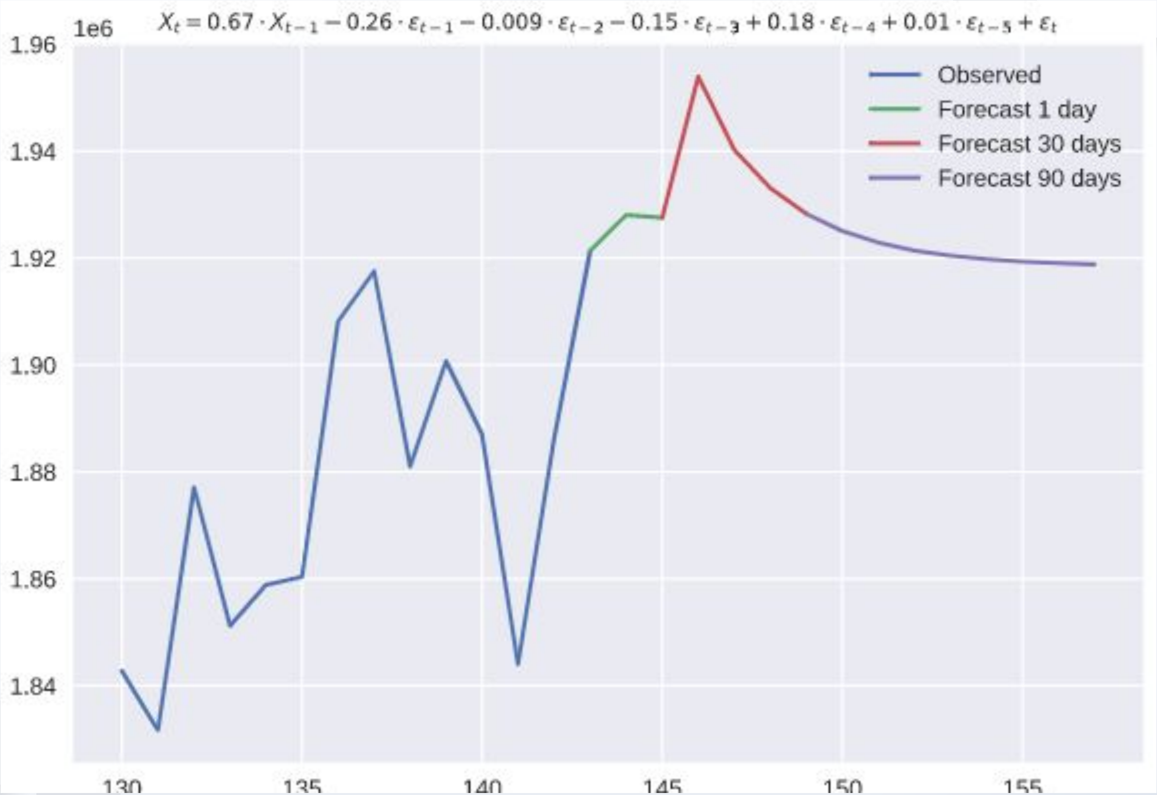


ARIMA Forecast Cluster 1



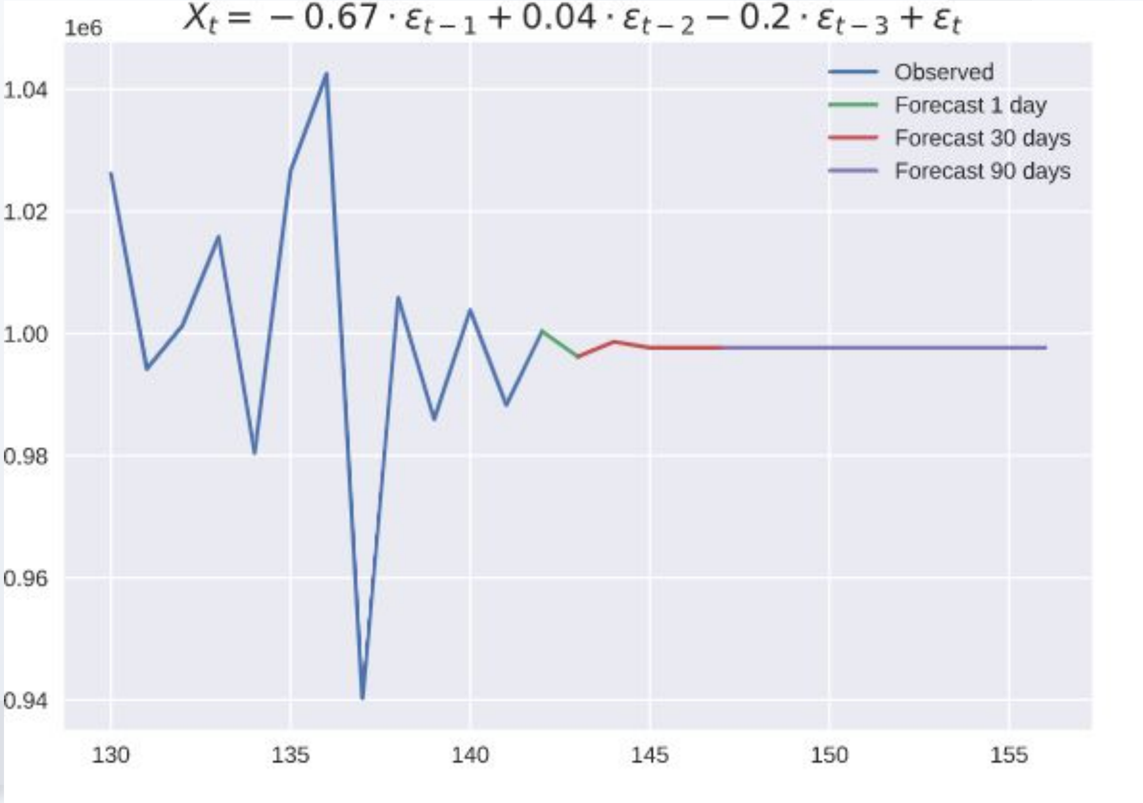
$$X_t = 0.87 \cdot X_{t-1} - 0.67 \cdot \varepsilon_{t-1} + \varepsilon_t$$

ARIMA Forecast Cluster 2



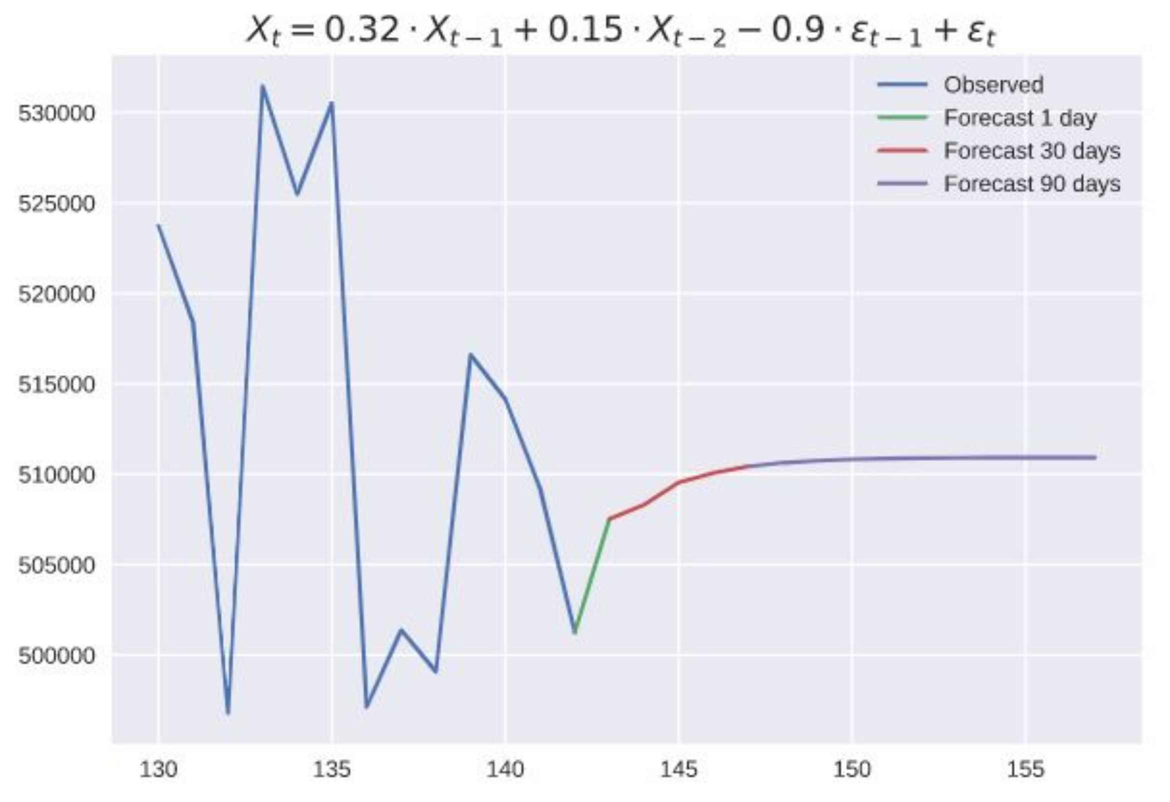
$$X_t = 0.67 \cdot X_{t-1} - 0.26 \cdot \varepsilon_{t-1} - 0.009 \cdot \varepsilon_{t-2} - 0.15 \cdot \varepsilon_{t-3} + 0.18 \cdot \varepsilon_{t-4} + 0.01 \cdot \varepsilon_{t-5} + \varepsilon_t$$

ARIMA Forecast Cluster 3

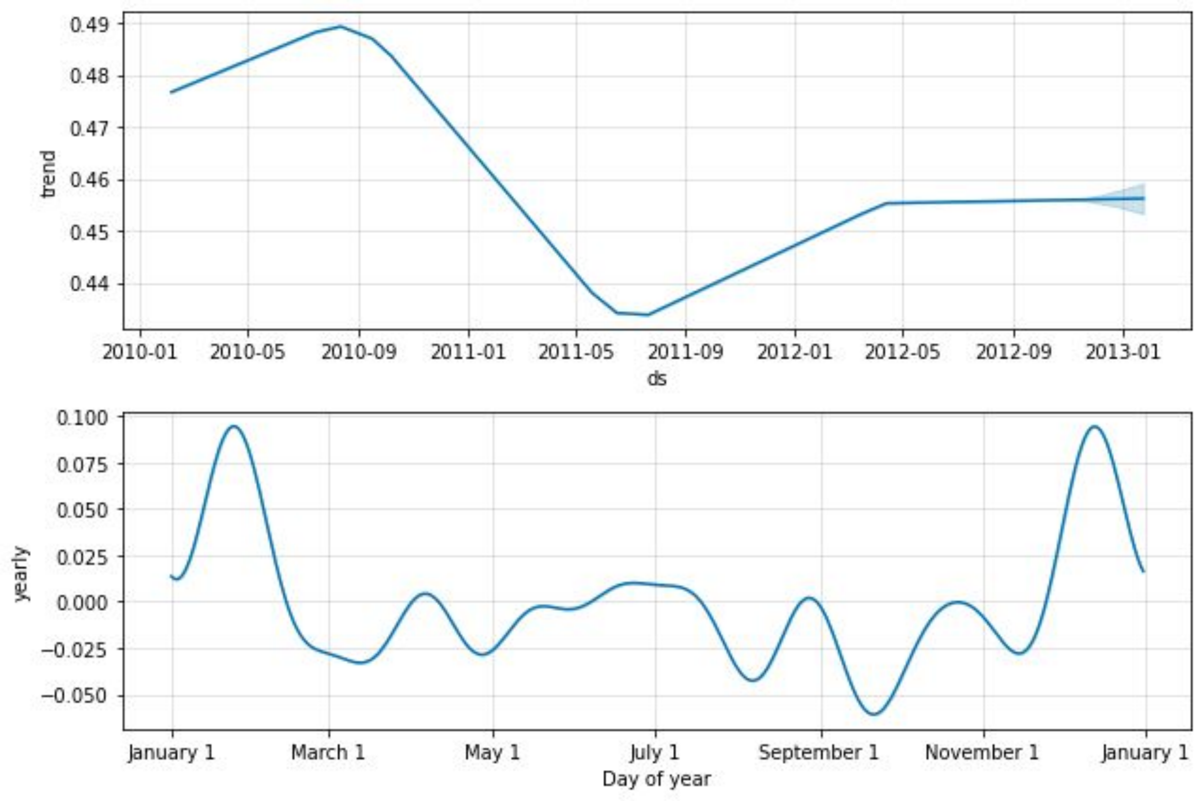


$$X_t = -0.67 \cdot \varepsilon_{t-1} + 0.04 \cdot \varepsilon_{t-2} - 0.2 \cdot \varepsilon_{t-3} + \varepsilon_t$$

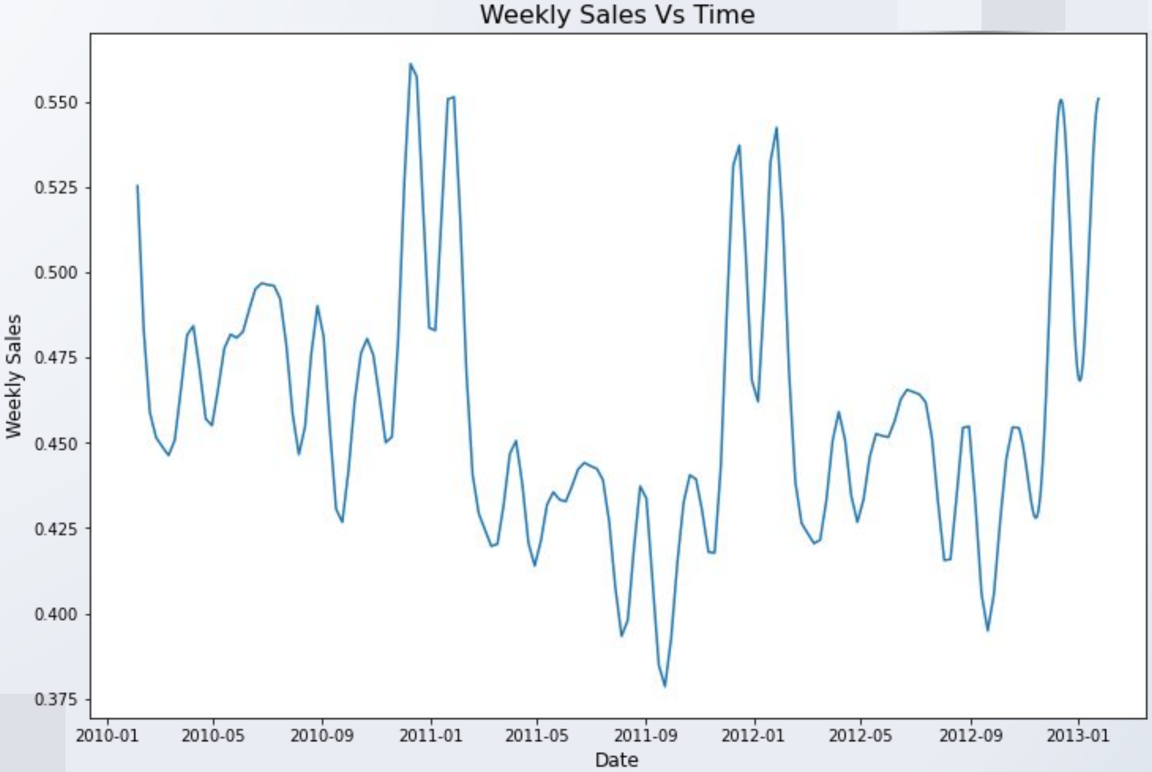
ARIMA Forecast Cluster 4



Facebook Prophet



Facebook Prophet



Conclusions

- ❑ The use of unsupervised methods allows for solving the problem with fewer models
- ❑ The application of multiple processing techniques like Elbow or normalization makes it easier to generate useful predictions

Thanks!

Any questions?