COMP 4254 Final Presentation

A Time Series Analysis with Temperature and Software Sales

By

Stan (Yuzhe) Chen

BCIT ID: A01042390

Introduction

- Objective

- We want to study if there is a relationship between temperature and sales of software in colder area of the world.

- We also want to do a time series forecast on the temperature and sales as a practice

- Data Sources

1) 1C Company sales, provided on Kaggle.

- It is an independent software developer, distributor and publisher with HQ in Moscow.

- Data set covers shops in multiple cities, different categories, online and offline, from 2013 to 2015.

2) Moscow Airport METAR (Meteorological Aerodrome Report) data, provided by rp5.ru

Data Preparation & Manipulation

- Sales data were filtered for shops in Moscow or the Online ones, then aggregated, resulted in daily and monthly items sold and revenue

- Temperature (in ^{o}C) data were aggregated to get average by day and month as well

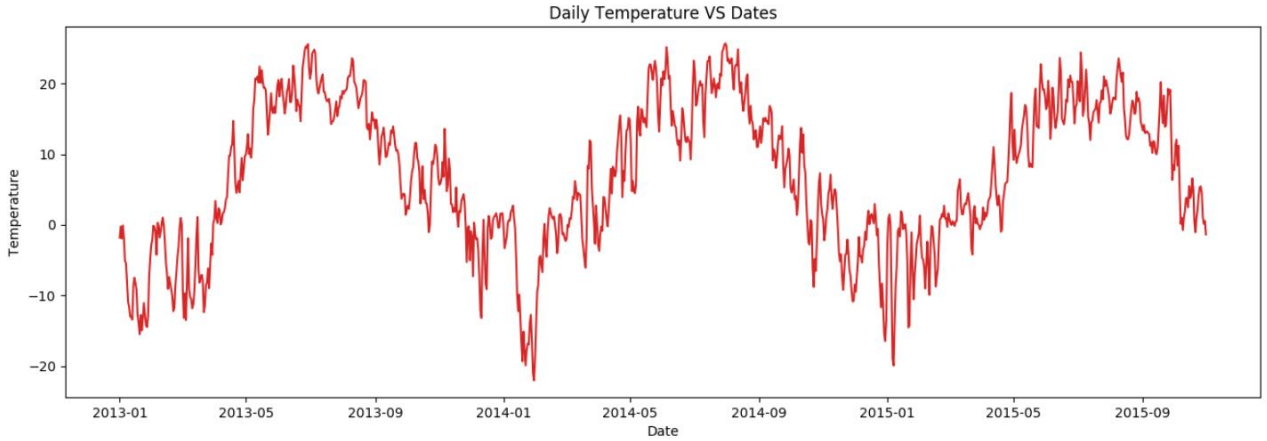
- The sales data and temperature were joint by having the same date for daily set, and by same month-year for monthly set

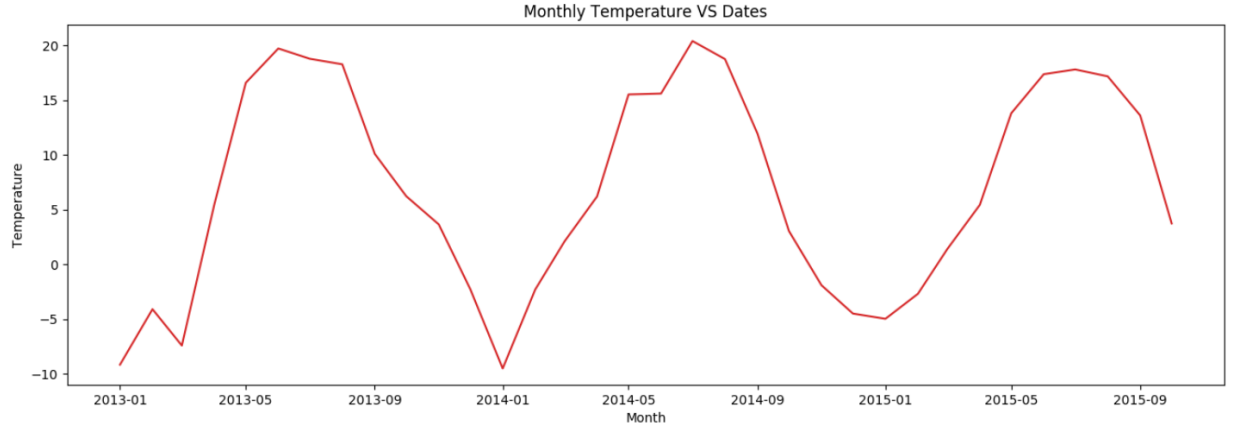
Explorative Analysis

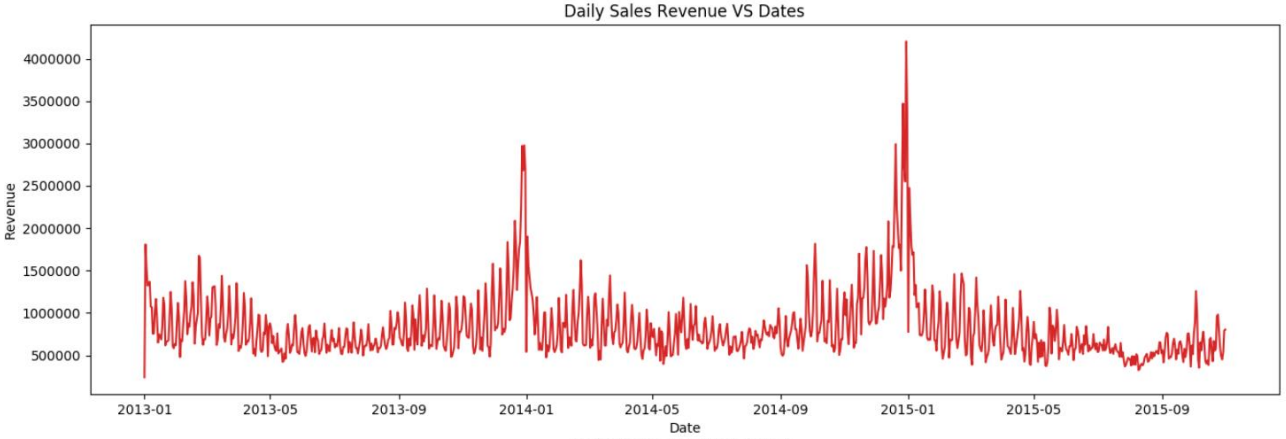
Revenue Contributor (201301~201510)	Million Rubs
Total 1C Sales of all	3398.93
Shops in Moscow Sales	1132.71
Online Sales	49.79

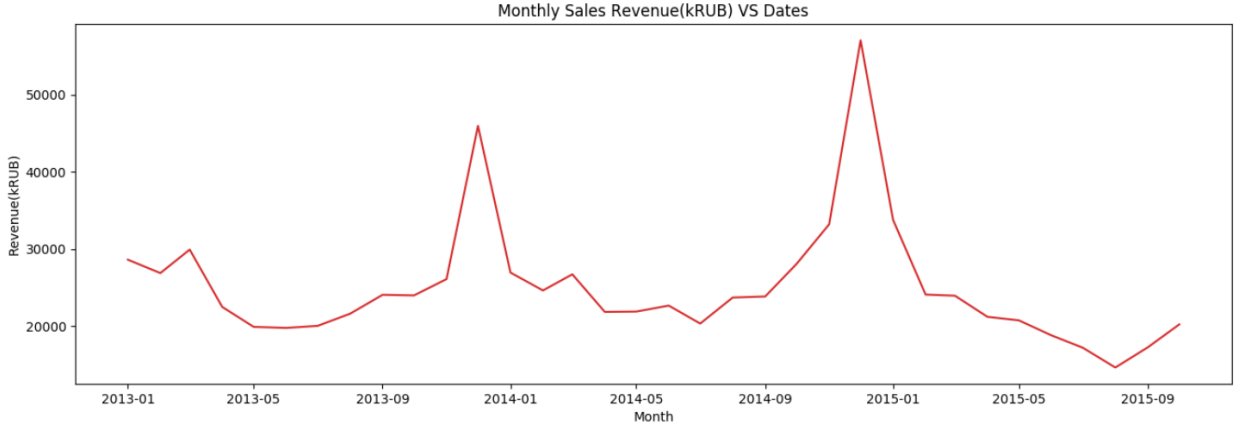
From above, total sales from 2013-01 to 2015-10 is about 3398.9 MRub, and revenue from shops in Moscow is around 1/3 of that, at 1132.7 MRub. Interestingly, the sales from Online is only about 49.8 MRub, not very significant contribution. In this case, we will just focus on the sales in Moscow.

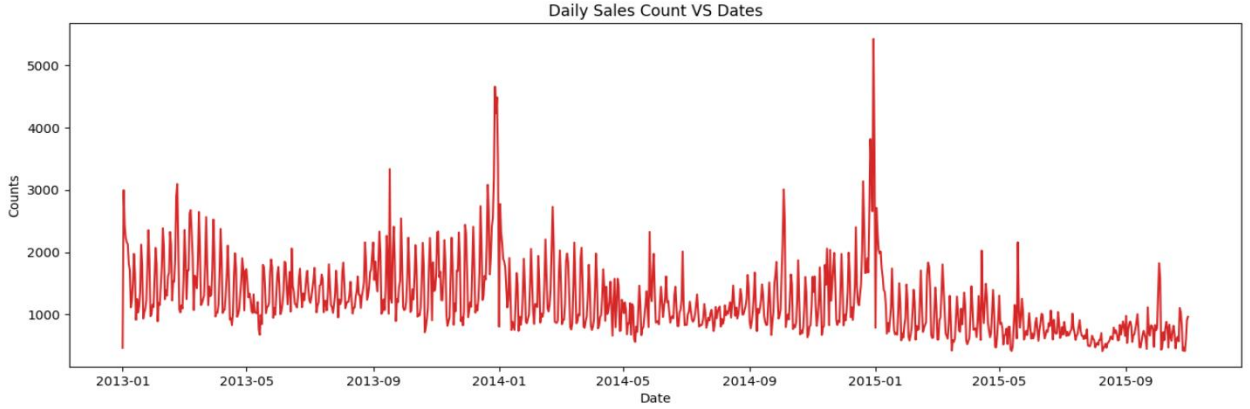
Visualization of Raw Data

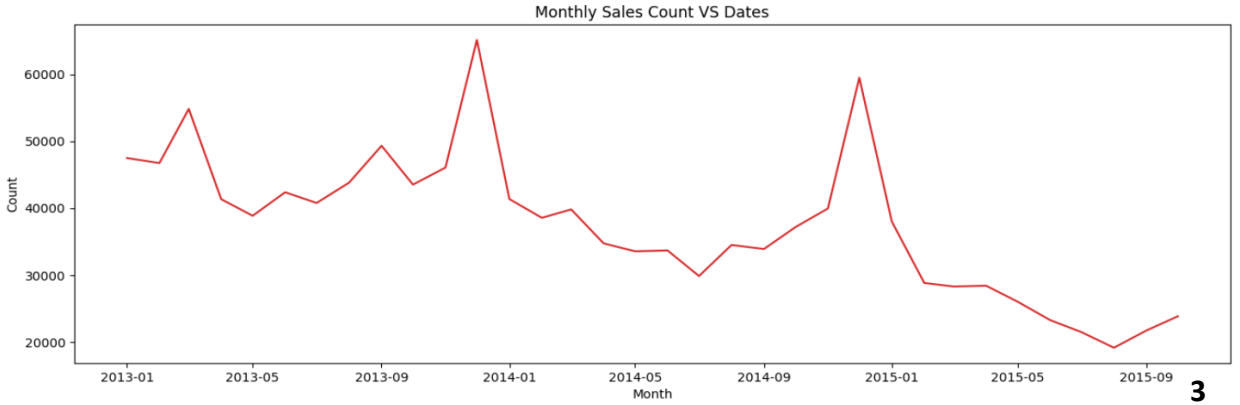












Some Statistical Tests…

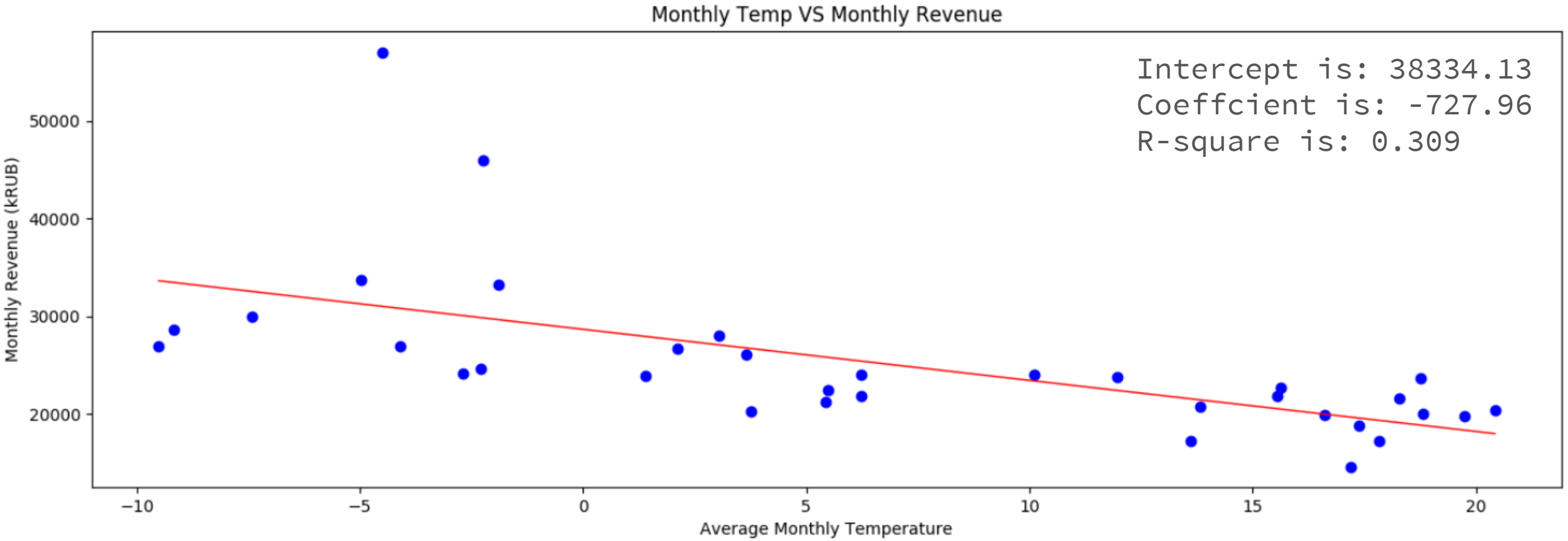
ADF Test
for stationarity,
with H_0 that time
series not stationary

Series	P -Value
Daily Average Temperature	0.05557
Daily Revenue	0.00051
Daily Items Count	0.00577
Monthly Average Temperature	7.3077e-06
Monthly Revenue (kRUB)	0.01704
Monthly Items Count	0.1681

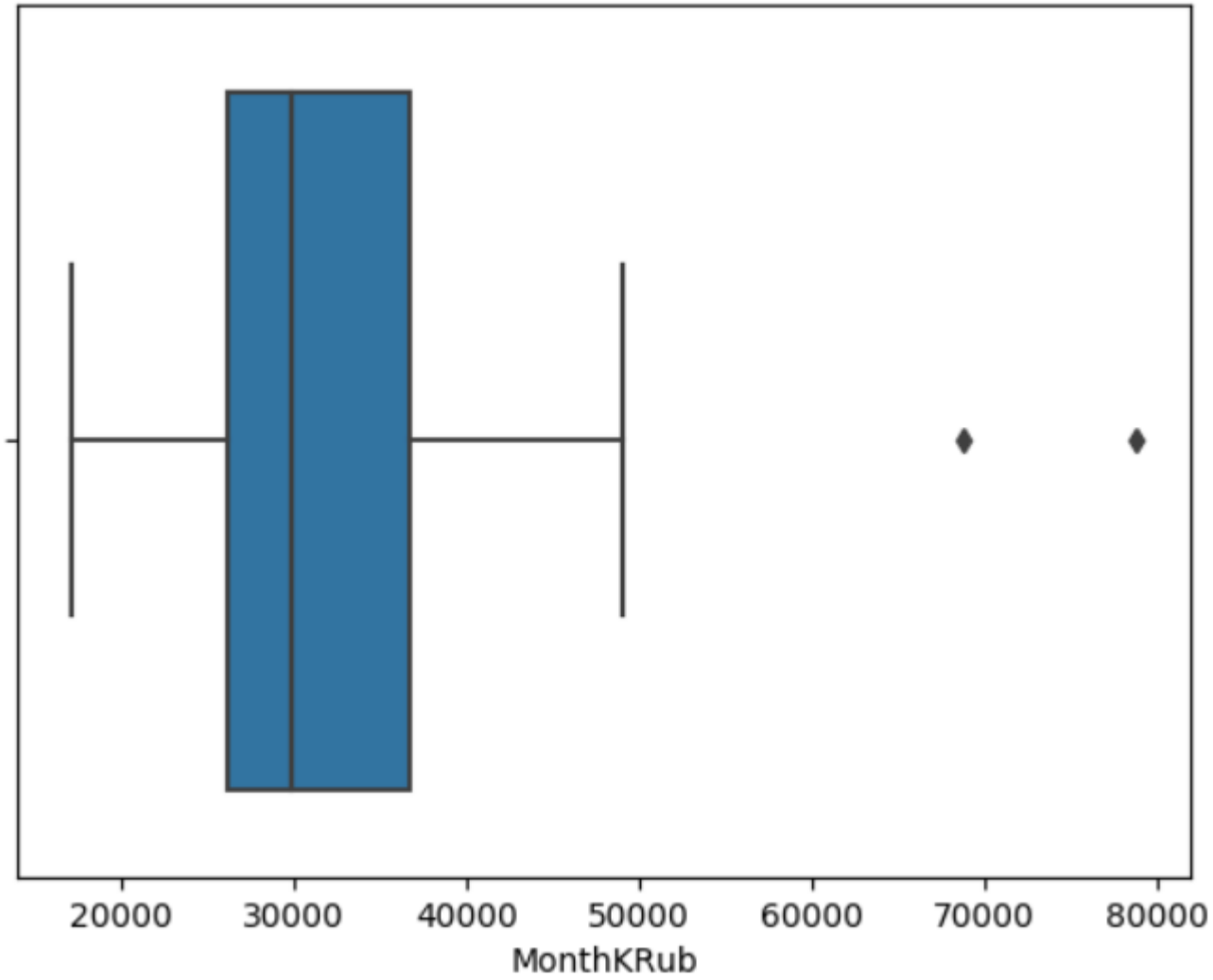
Pearson's Correlation
Coefficient,
absolute value closer to 1
means perfect linearity

Pairs of Variables	Pearson's Correlation Coefficient
Daily Average Temperature VS Revenue	-0.284
Daily Average Temperature VS Items Count	-0.283
Monthly Average Temperature VS Revenue (kRUB)	-0.556
Monthly Average Temperature VS Items Count	-0.490

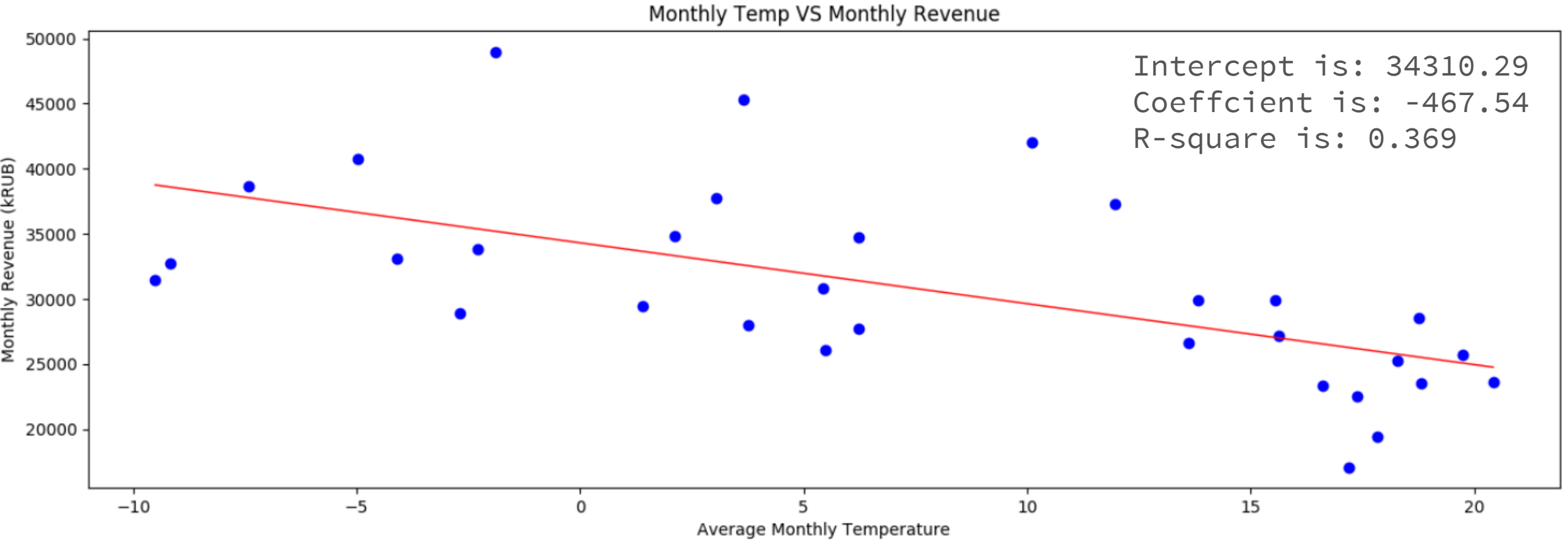
Since pair of Monthly Average Temperature VS Revenue has largest Pearson's Correlation Coefficient in absolution value, being most promising, and we will focus on them.



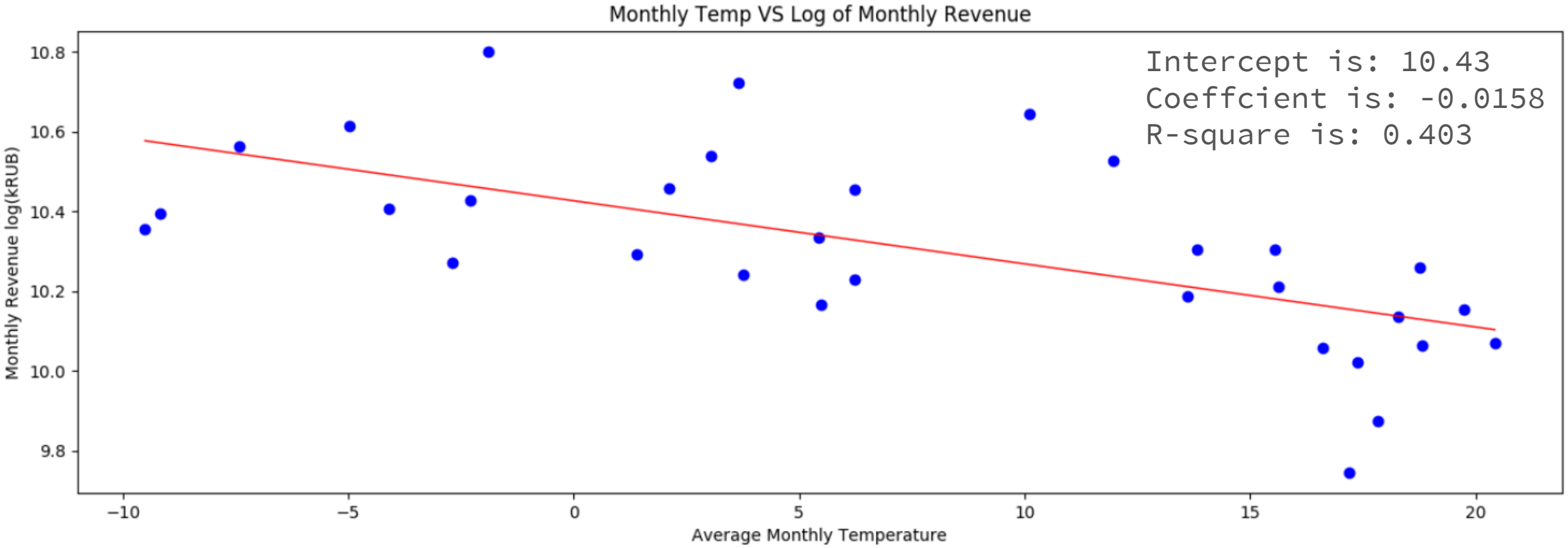
We can try to improve the Monthly Average Temperature VS Revenue model by removing the outliers first, which can be detected using the box-plot.



Then we do a new fitting for the Monthly Average Temperature VS Revenue model, which improves the R-square value to 0.369.

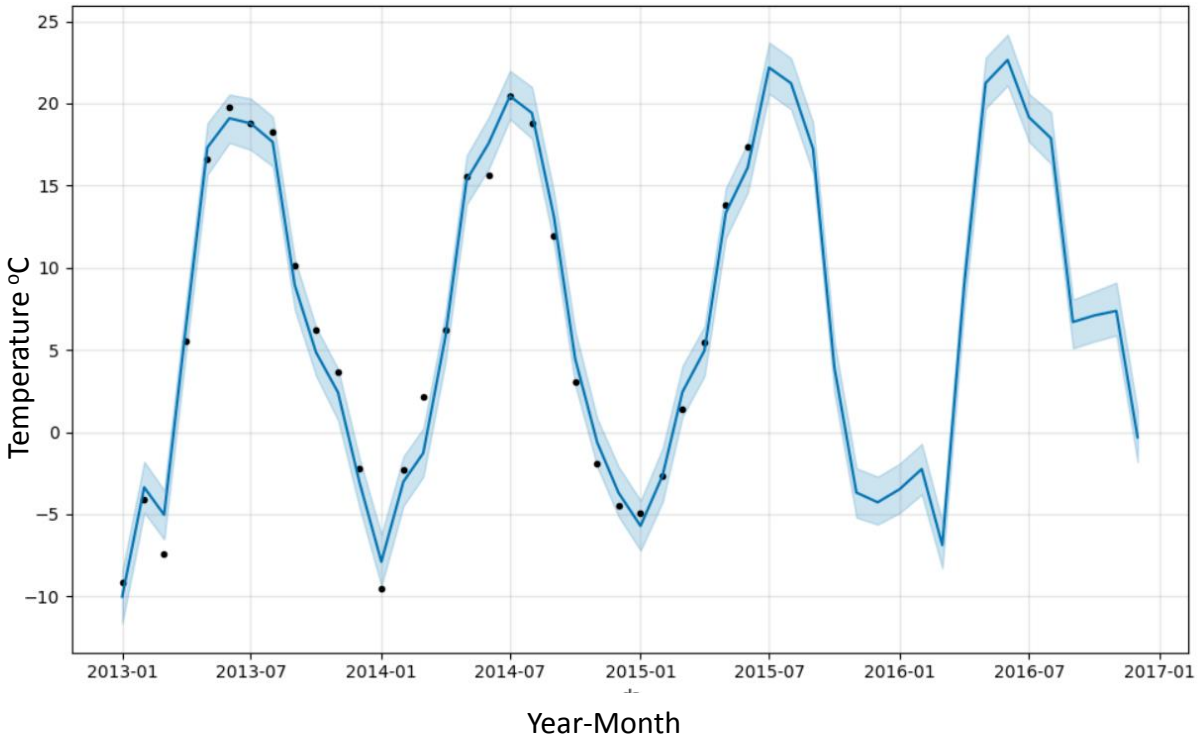


We can further improving the Monthly Average Temperature VS Revenue model, by taking a log for the monthly revenue in kRUB. This improves the R-square value to 0.403.

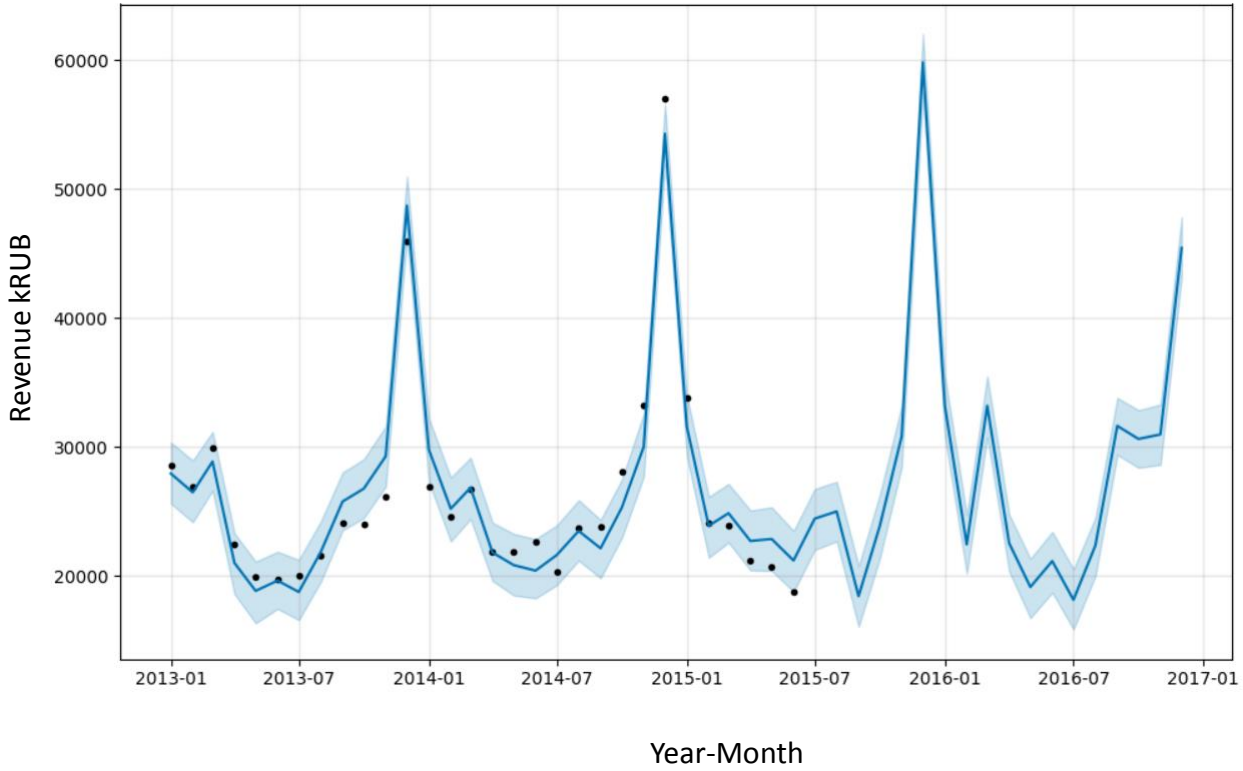


We also do a time series forecast of the two, with Facebook Prophet

Training Set: 2013-01 ~ 2015-06; Validation/test Set: 2015-07 ~ 2015-11



Mean: 6.89
Standard Deviation: 9.57
RMSE: 3.47

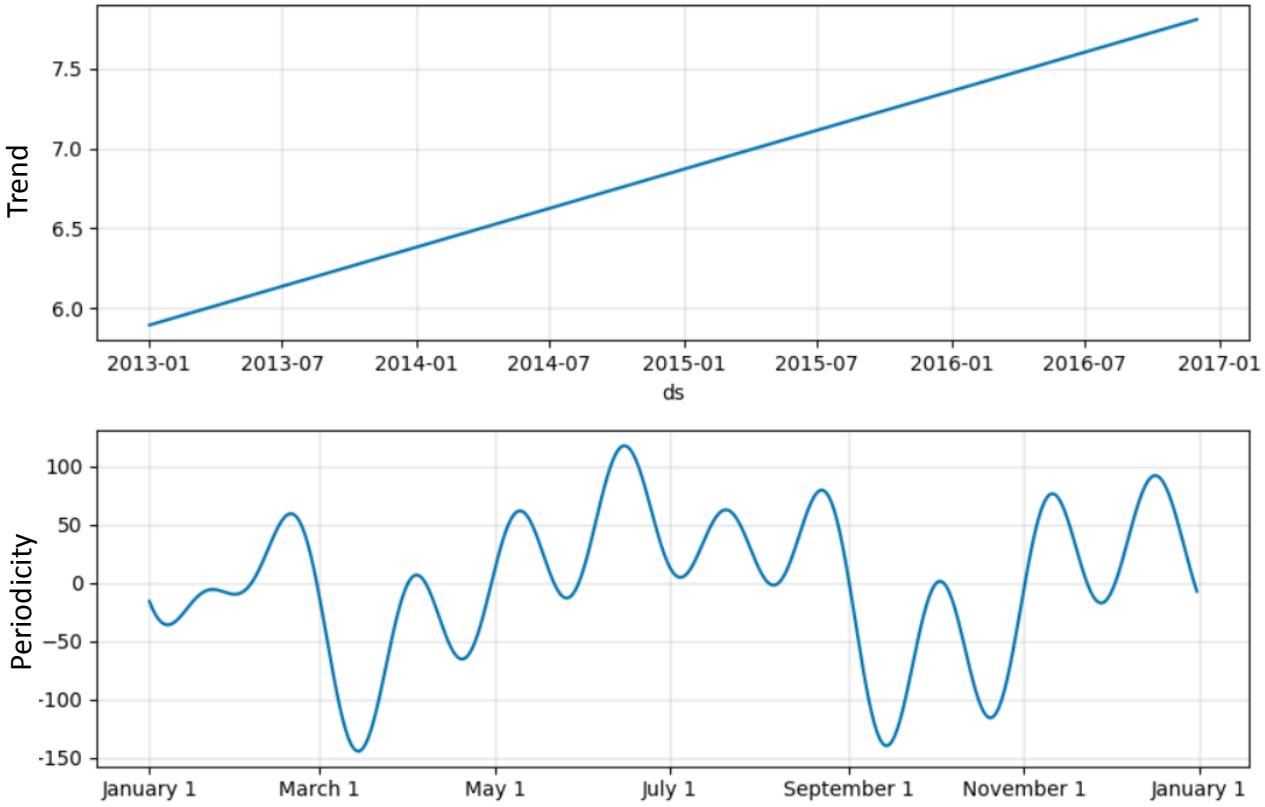


Mean: 33313.89
Standard Deviation: 10755.10
RMSE: 6,118.02

More information on Facebook Prophet:

- It is fast and provides completely automated forecast that can be tuned by hand
- It is based on an additive model where non-linear trends are fit with yearly, weekly, and daily seasonality, plus holiday effects

Monthly Temperature (°C)



Monthly Revenue (kRUB)