Project Proposal

Solonas Kyprou

skyprou@gmail.com

The Python Quants — September 2, 2019

Title

The proposed project title is "Modeling mall customer visitation data using a Gaussian Process".

Abstract

1   Main Goal

The main goal of this project is to model the Thasos Mall REIT Foot Traffic Index using a Gaussian Process (GP). Afterwards I will briefly explore the possibility of using this model to constuct a signal for some relevant REITs by trying different metrics such as YoY change in visitation.

2   Thasos Group

Thasos is an award-winning alternative data intelligence firm based in New York City. Founded in 2011 at MIT, Thasos transforms real-time location data from mobile phones into objective and actionable information on the performance of businesses, markets, and economies worldwide.

Thasos MallStreams products enable detailed near real-time analysis of over 95% of the malls owned or operated by the largest publicly traded retail Real Estate Investment Trusts (REITs) such as GGP (Brookfield Properties Retail Group) and SKT (Tanger Factory Outlet Centers Inc.).

The Thasos Mall REIT Foot traffic Index is seasonal with peaks on Black Friday and during the Christmas holiday period. It is expexted that this seasonality can be modelled using a Gaussian Process. The figure below shows the customer visitation data for the properties owned by 4 REITs. It is shown just to demonstrate the shape of the data that will be modelled. The actual data of the project will just be one aggregate time series of all REITs, which is available on Bloomberg for free.
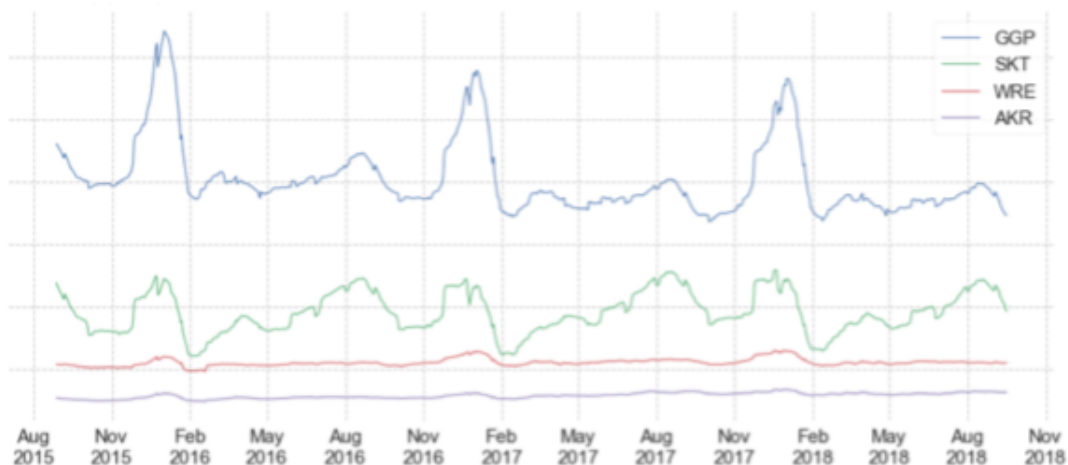
3 Gaussian Process (GP)

In probability theory and statistics, a Gaussian process is a stochastic process (a collection of random variables indexed by time or space), such that every finite collection of those random variables has a multivariate normal distribution, i.e. every finite linear combination of them is normally distributed. The distribution of a Gaussian process is the joint distribution of all those (infinitely many) random variables, and as such, it is a distribution over functions with a continuous domain, e.g. time or space.

A machine-learning algorithm that involves a Gaussian process uses lazy learning and a measure of the similarity between points (the kernel function) to predict the value for an unseen point from training data. The prediction is not just an estimate for that point, but also has uncertainty information—it is a one-dimensional Gaussian distribution (which is the marginal distribution at that point).

4 Python packages

Standard python packages are expected to be used, such as Pandas, Numpy, TensorFlow and scikit-learn.

5 Hypothesis/Expectations

The hypothesis is that we can predict future customer visitations (time-horizons will be decided at a later stage) by building a GP model on historical data. Then I will briefly explore whether there are any alpha opportunities on REITS by using this model. It is expected that the Gaussian Process model will have a good statistical power, however it is not expected that there will be any strong alpha opportunities as there are many other significant factors affecting the stock prices that are out of scope of this project.