

PGGM Quant Case

Case Study – Systematic Equity Strategies

Maart 2023

Prerequisites + Installation guide Python & R

Introductie

Voor de Quant Case tijdens de business course is het nodig om te beschikken over software waarmee je kan modelleren. Hieronder vind je hoe je zulke software kan installeren, mocht je hier nog niet over beschikken. Python is de meest gebruikte programmeertaal voor dit type analyse binnen PGGM. Hoewel we het in dit document houden op Python en R, is het natuurlijk ook mogelijk om andere software te gebruiken. Echter raden wij het erg aan om gebruik te maken van Python en dit toe te passen binnen een Anaconda Environment. Zorg dat je dit op orde hebt voorafgaand aan de business course, dan kun je je gedurende de case volledig focussen op de opdracht.

Neem mee:

1. **Een eigen laptop**
2. **Met daarop geïnstalleerd: een werkende Python interpreter of RStudio.**
 - a. Python: in ieder geval de packages *pandas*, *numpy*, *statsmodels* en *sklearn*
 - b. In het geval van R: *Caret*, *Ggplot2*, *randomForest* en *gbm*

Python – Installatie

Python wordt veel gebruikt binnen PGGM, onder andere voor modelleren, data management en dashboarding. De meest gebruikte programma's voor het werken met Python te werken zijn Jupyter Notebook, Spyder, PyCharm en VScode. Een gemakkelijke manier om toegang te krijgen tot meerdere programma's is door Anaconda te downloaden: een package- en environment manager tegelijk.

Installatie Anaconda

1. Navigeer naar <https://docs.anaconda.com/anaconda/install/> en selecteer de juiste installatie op basis van je OS.
2. Volg de installatie manual en installeer Anaconda. Mocht je gebruik willen maken van Pycharm, dan kan dat aan het einde van de installatie.
3. Zodra Anaconda geïnstalleerd is kun je via de "Anaconda Navigator" Python applicaties zoals Jupyter Notebooks en Spyder openen.

Python – Aangeraden packages

Wij raden aan de volgende packages van te voren te installeren;

- NumPy – Arrays
- Pandas – Dataframes
- Scikit-learn (sklearn) – Machine learning
- Statsmodels – Statistische analyse
- Matplotlib – Visualisatie

Package installatie

Voordat je een package kan gebruiken in Python moet je deze eerst installeren. Dit kan zowel in de Python Terminal of in de Anaconda Prompt.

Python Terminal

Het installeren van packages in de Python Terminal gaat met het command:

pip install "PACKAGE"

Om bijvoorbeeld Pandas te installeren type je: *pip install pandas*

Anaconda Prompt

Mocht je Anaconda Prompt gebruiken om packages te installeren dan is de code net anders:

conda install "PACKAGE"

Package Importeren in je code

Nadat je een package geïnstalleerd hebt, moet je hem importeren in je script om te kunnen gebruiken. Dat gaat met de volgende code:

Import "PACKAGE" as "SHORTNAME"

Bijvoorbeeld: *import pandas as pd* of *import scikit-learn as sk*.

Let op! Iedere keer dat je een command gebruikt van een geïmporteerde package, moet je de "shortname" gebruiken. Bijvoorbeeld: *pd.DataFrame* voor de DataFrame command van pandas.

R – Installatie

R wordt met name gebruikt voor statistische analyses binnen PGGM en het meest gebruikte programma is RStudio.

Installatie R Studio

1. Navigeer naar <https://cran.r-project.org/bin/windows/base/> en installeer de laatste versie van R. Deze stap is vereist omdat R Studio hierop voort bouwt. Voor Mac zie: <https://cran.r-project.org/bin/macosx/>
2. Ga vervolgens naar <https://rstudio.com/products/rstudio/download/> en installeer R Studio (gratis versie).
3. Volg de instructies vanuit R studio.

R – Aangeraden packages

Voor R kan je de volgende packages eventueel installeren:

- Caret – Machine learning
- Ggplot2 – Visualisatie
- randomForest – Machine learning (Random Forests)
- gbm – Machine learning (Gradiënt Boosting)

In R is het installeren en gebruiken van packages eenvoudig. Voor het installeren van een package gebruik je de volgende code:

```
install.packages("PACKAGE")
```

Voor ggplot2 is dit bijvoorbeeld: *install.packages("ggplot2")*

Na installatie kun je gebruik maken van de package door de volgende command te gebruiken:

```
library("PACKAGE")
```

Voor ggplot2 is dit bijvoorbeeld: *library(ggplot2)*. Na import kun je de commands vanuit de nieuwe packages vrij gebruiken.