**Topos**

**Data Engineer Intern Assignment**

**By**

**Ronak Jayantilal Parakhiya**

**Version: 1.0**

**Date: 05/28/2019**

Table of Contents

[Python Installation 2](#_Toc9957767)

[Jupyter Installation 3](#_Toc9957768)

[Pre-requisite 3](#_Toc9957769)

[Installation using Anaconda 3](#_Toc9957770)

[Installing with pip 3](#_Toc9957771)

[Data Analysis Libraries 4](#_Toc9957772)

[Pandas 4](#_Toc9957773)

[Description 4](#_Toc9957774)

[Install Panda 4](#_Toc9957775)

[Beautiful Soap 4](#_Toc9957776)

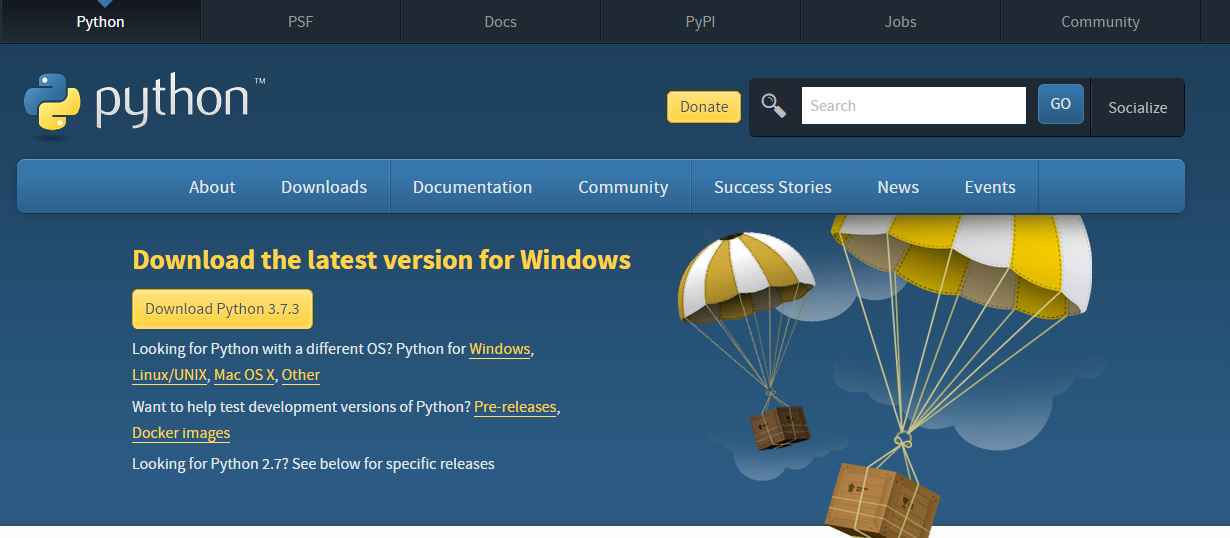
[Description 4](#_Toc9957777)

[Install Beautiful Soap 4](#_Toc9957778)

# Python Installation

Below are the steps to install Python to run the scrapper

1. Click on link to go to official website of python: <https://www.python.org/downloads/>
2. Click on the download button of latest version as shown in below screenshot



1. After successful download please accept the license agreement and click on next to continue
2. Click on “Finish” for successful installation of python

# Jupyter Installation

## Pre-requisite

Python is a requirement (Python 3.3 or greater, or Python 2.7) for installing the Jupyter Notebook itself

## Installation using Anaconda

First, download Anaconda. We recommend downloading Anaconda’s latest Python 3 version.

Second, install the version of Anaconda which you downloaded, following the instructions on the download page.

To run the notebook, run the following command at the Terminal (Mac/Linux) or Command Prompt (Windows):

* jupyter notebook

## Installing with pip

If you have Python 3 installed (which is recommended):

* python3 -m pip install --upgrade pip
* python3 -m pip install jupyter

If you have Python 2 installed:

* python -m pip install --upgrade pip
* python -m pip install jupyter

# Data Analysis Libraries

## Pandas

### Description

Pandas is an open source, BSD-licensed library providing high-performance, easy-to-use data structures and data analysis tools for the [Python](https://www.python.org/) programming language.

### Install Panda

1. To install using conda-forge execute below command

* Conda install panda

1. To install via PyPI

* python3 -m pip install --upgrade pandas

## Beautiful Soap

### Description

Beautiful Soup is a Python library for pulling data out of HTML and XML files. It works with your favourite parser to provide idiomatic ways of navigating, searching, and modifying the parse tree.

### Install Beautiful Soap

For Windows use below command

* pip install beautifulsoup4

For Ubuntu Linux use below command

* apt-get install python-bs4