DATASCI 410

Data Science: Methods for Data Analysis

Week 1

PROFESSIONAL & CONTINUING EDUCATION
UNIVERSITY of WASHINGTON

About this Course

1. This course includes a lot of material.
2. Expect and plan to spend several hours outside the classroom on this course.
3. This course is intense. Feeling frustrated at times is ok.
4. Keep up with assignments, milestones, discussions. There will be no time for catching up.
5. Did I mention this course covers a lot of material?

Course Objectives

1. Apply the data science process to a business problem including determining data requirements, exploring the data, and presenting actionable results and recommendations.
2. Explore complex data relationships and present results in an insightful manner to a non-technical audience.
3. Apply basic concepts of probability and statistics including conditional probability, sampling, and hypothesis testing.
4. Generalize the theory and practice of linear models as a foundation for machine learning.
5. Apply basic time series models for forecasting, simple text analytics, and unstructured data analysis

Grading Components

Course is graded as Satisfactory (>80%) or Unsatisfactory.

Component	Percentage
Participation	16%
Quizzes	20%
Lesson Assignments	20%
Milestone Projects 1, 2 and 3	24%
Milestone Project 4	20%

Course Assignments & Due Dates

Weekly assignments have 2 weeks

- After the first week it is marked late.
- After the 2nd week, it is unavailable to submit.

3 types of assignments to submit:

- Lesson Jupyter Notebooks
- Lesson Opinion Editorials
- Milestone Jupyter Notebooks
  - Look ahead at the Milestone Overviews

Course Topics

Part 1 – Data Visualization	Part 2 – Statistical Analysis	Part 3 – Linear Models	Part 4 - Other Machine Learning Models
Lesson 1 – Data Exploration Part 1 Lesson 2 – Data Exploration Part 2	Lesson 3 – Combinations, Permutations, & Probability Lesson 4 – Sampling & Hypothesis Testing Lesson 5 – Introduction to Bayes Theorem	Lesson 6 – Introduction to Regression Lesson 7 – Regression & Regularization Lesson 8 – Time Series Analysis	Lesson 9 – Näive Bayes Lesson 10 – Basic Text Analytics
Milestone 1 – Data Visualization Complementary Views	Milestone 2 – Hypothesis Simulation	Milestone 3 – Regression Models	Milestone 4 - Independent Project

Translates into:

1. Learn methods to explore and understand data.
2. Understand the core concepts of statistics and probability.
3. Understand and implement various statistical procedures in Python.
4. Understanding the mathematical basis of machine learning models.
5. Expand Python programming skills to be able to write and test quality code from scratch.

Course Technology

Python 3

Anaconda Distribution

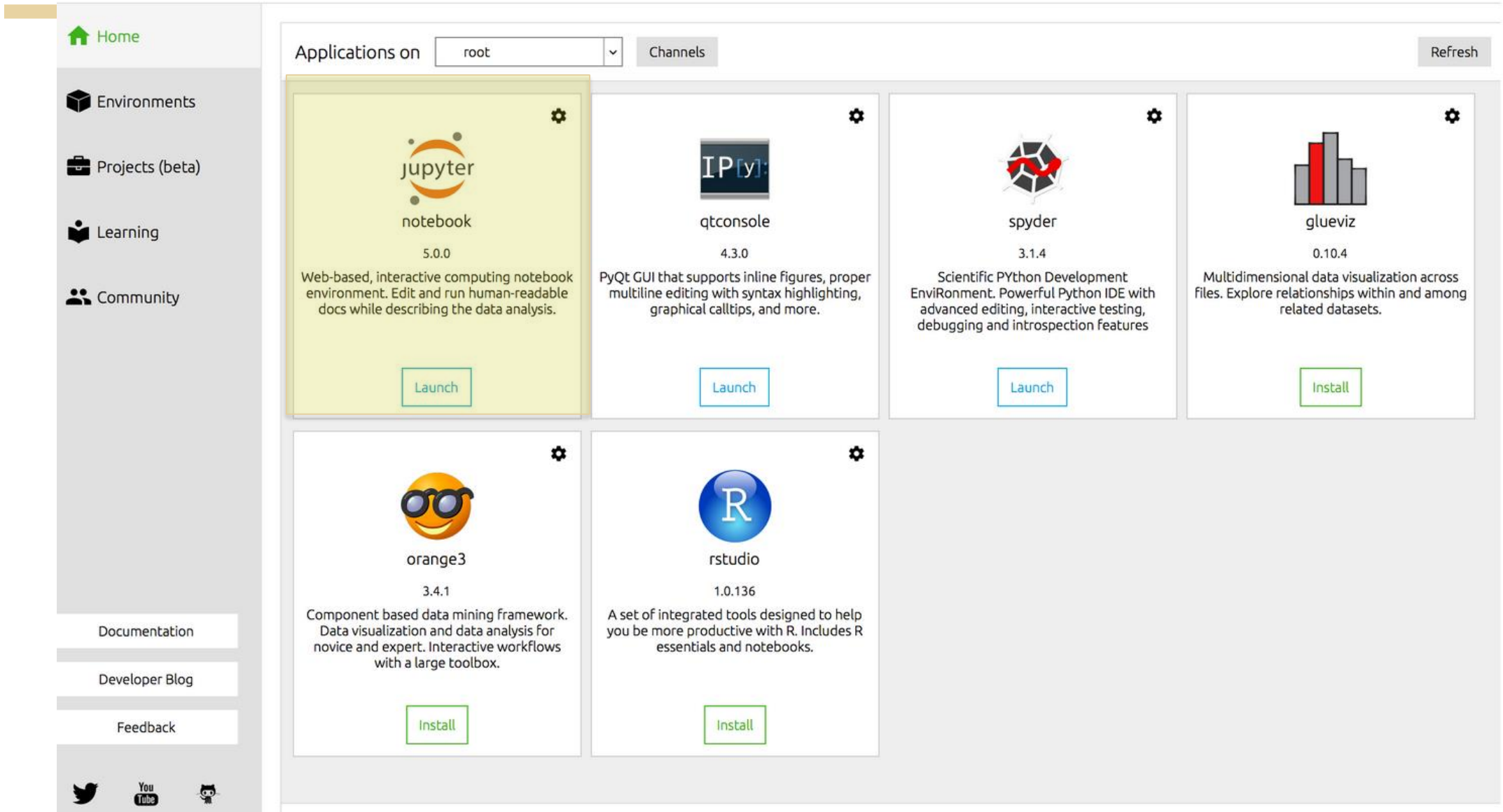
- Jupyter Notebook

Course Technology

Introduction to Jupyter notebook:

https://medium.com/codingthesmartway-com-blog/getting-started-with-jupyter-notebook-for-python-4e7082bd5d46

Anaconda Navigator



Update your packages

conda update <packagename>

- conda
- setuptools
- python
- pandas
- numpy
- notebook
- seaborn
- matplotlib
- ipython

Accessing Virtual Labs

If you cannot install Jupyter Notebook on your machine

From the Labs link in the Lesson

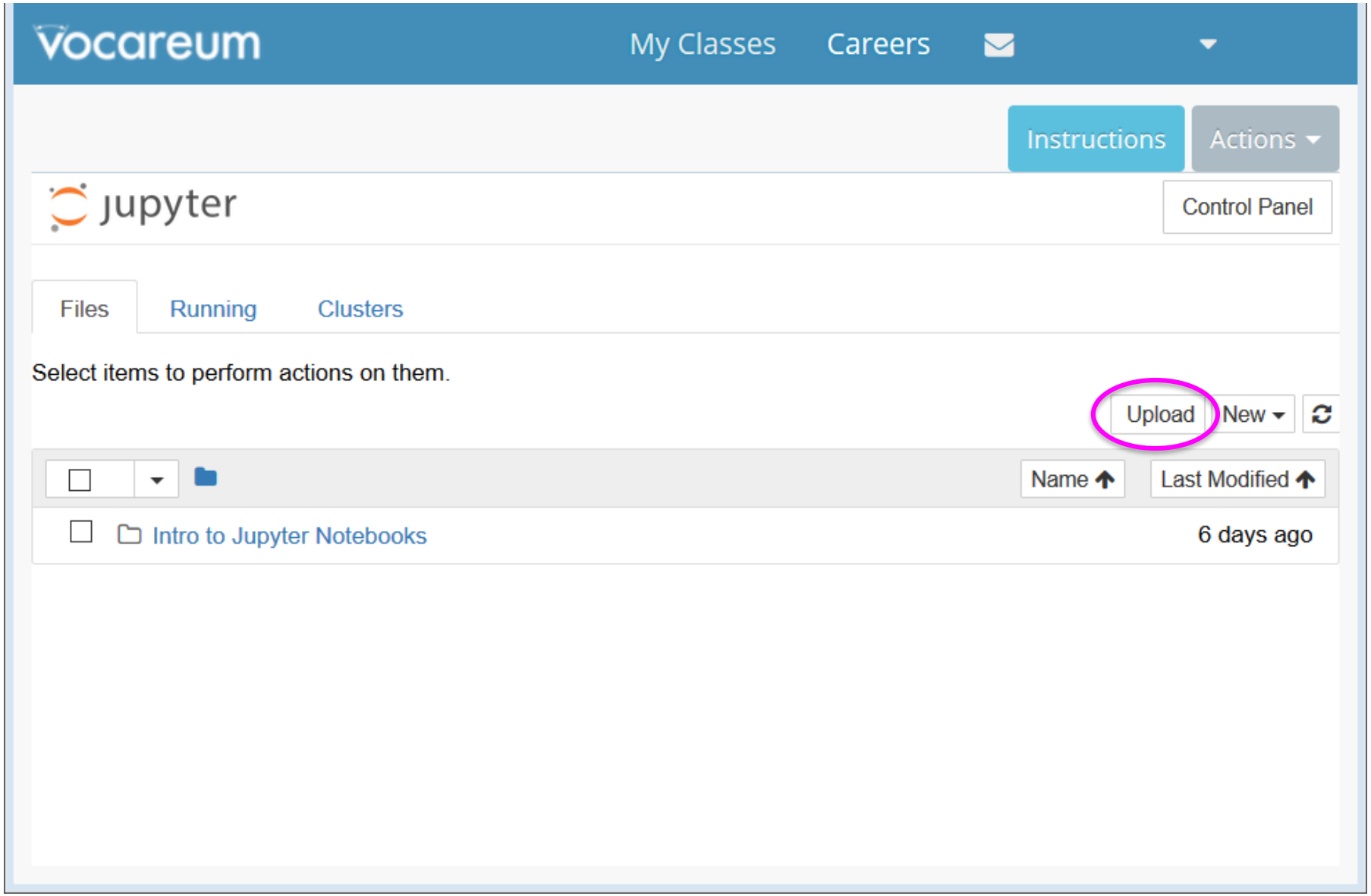
Virtual Lab will launch in a new window.

Load Vocareum Virtual Lab in a new window

Every lesson has Juypter Notebook lab exercises.

- Upload the Notebook to your virtual lab

Vocareum Virtual Lab



Summary

- >Attend class every week
  - –Work on your own in addition to class night
- >Install Anaconda Distribution with Jupyter Notebook
- >Get comfortable with Data Munging
  - –Data exploration
  - –Statistical Analysis
  - –Building & Improving Models