Data Science with Python

1. <u>Course overview</u>

Data science is an interdisciplinary field that gains insights into data through computation, statistics and visualization. This course introduces fundamental machine learning algorithms and techniques for analyzing and visualizing real-world data using Python-based programming tools. Topics include tutorial in Python, Jupyter Notebook, Python libraries and packages, exploratory data analysis, machine learning algorithms, data visualization, case studies of real-world data sets as well as introduction to advanced topics in machine learning and artificial intelligence. This course includes a hands-on data science project in which students apply techniques and skills in statistical data analysis, machine learning methods and visualization to solve real-world data science problem. After completing this course, students will be able to perform basic data manipulation and apply machine learning algorithms to solve real-world data science problem.

2. <u>Pre-requisites</u>

Prior programming experience is preferred but not required. Students must bring a laptop computer to each session.

3. <u>Course instructor</u>

Dr. X. Wang holds a PhD in Astrophysics from Harvard University. She has over ten first-authored publications in top tier scientific journals including *Nature Physics*. She was a faculty member at Harvard Summer School teaching the Pre-College Program in Data Science and Astronomy. Dr. Wang is very interested in teaching science and making data science accessible and explainable to a broader audience.

4. <u>Course logistics</u>

This course includes 10 weekly one-hour sessions offered virtually via zoom.
Minimum cohort size: 4
Date: 11/08/2020-01/10/2021, 9-10am every Sunday (Beijing time).
Tuition: $500
Cancellation policy: any cancellation should be made *before* the second session and will receive a refund of $450.

5. <u>Course expectations</u>

Participation. The sessions will be conducted in an active, collaborative learning environment. Rather than a traditional lecture-based style, students will experience active learning exercises where they discuss and work in groups and present their work at the end of presentation of new concepts. These in-session hands-on coding exercises are aimed to deepen the understanding of the material and help to develop analytical thinking and problem-solving skills.

Project. Student will work on a data science project in groups or on an individual basis. Students are encouraged to discuss and refine their project ideas with the lecturer. Students will acquire data, leverage machine learning algorithms, design visualization and present their results. Each group/individual will give a 15-minute presentation followed with a 5-minute Q&A in the last session (format subject to cohort size and background).

6. <u>Course outline (tentative)</u>

Date	*Session topic*	*Content & goal*	
11/08/2020	Introduction to data science	-	Course intro
		-	Introduction to data science: what is it, how does it work and why we need it
11/15/2020	Python tutorial	-	Software installation
		-	Python & Jupyter notebook tutorial
		-	Hands-on coding exercise
11/22/2020	Exploratory data analysis I	-	Data cleaning
		-	Visualization tools in Python
		-	Basic statistics
11/29/2020	Exploratory data analysis II	-	How to dig insights from data
		-	EDA
12/06/2020	Introduction to machine learning	-	Introduction to machine learning
		-	Common algorithms and techniques
		-	Data science pipeline & framework
12/13/2020	Linear and logistic regression	-	Fundamentals of regression techniques
		-	Implementation of linear and logistic regression in Python
		-	Hands-on exercise
12/20/2020	Tree-based algorithms	-	Introduction of tree-based machine learning algorithms (e.g. decision tree, random forest)
		-	Implementation in Python
12/27/2020	Real-world data science problem I	-	Hands-on demo of a real-world data science problem I
01/03/2020	Real-world data science problem II	-	Hands-on demo of a real-world data science problem II
01/10/2020	Final presentation	-	15-min presentation followed by a 5-min Q&A
		-	Course retrospective
		-	Future learning opportunities