**Exploratory Project**

**20 MINUTES**

Project: **Titanic Survival Exploration**

In this ***optional*** project, you will create decision functions that attempt to predict survival outcomes from the 1912 Titanic disaster based on each passenger’s features, such as sex and age. You will start with a simple algorithm and increase its complexity until you are able to accurately predict the outcomes for at least 80% of the passengers in the provided data. This project will introduce you to some of the concepts of machine learning as you start the Nanodegree program.

**1 LESSON, 1 PROJECT**

**Project Overview**

Welcome to the Machine Learning Engineer Nanodegree!

In this ***optional*** project, you will create decision functions that attempt to predict survival outcomes from the 1912 Titanic disaster based on each passenger’s features, such as sex and age. You will start with a simple algorithm and increase its complexity until you are able to accurately predict the outcomes for at least 80% of the passengers in the provided data. This project will introduce you to some of the concepts of machine learning as you start the Nanodegree program.

In addition, you'll make sure Python is installed with the necessary packages to complete this project. There are two Python libraries, numpy and pandas, that we'll use a bit here in this project. Don't worry about how they work for now — we'll get to them in Project 1. This project will also familiarize you with the submission process for the projects that you will be completing as part of the Nanodegree program.

**Software and Libraries**

This project uses the following software and Python libraries:

* [**Python 2.7**](https://www.python.org/download/releases/2.7/)
* [**NumPy**](http://www.numpy.org/)
* [**pandas**](http://pandas.pydata.org/)
* [**matplotlib**](http://matplotlib.org/)
* [**iPython Notebook**](http://ipython.org/notebook.html)

If you already have Python 2.7 installed on your computer, then you can install NumPy scikit-learn, and iPython Notebook by using [**pip**](https://pip.pypa.io/en/stable/) on the command line. [**This page**](http://www.lfd.uci.edu/~gohlke/pythonlibs/) may also be of use for some packages for Windows users, if pip has trouble performing the installation.

If you do not have Python installed yet, it is highly recommended that you install the [**Anaconda**](http://continuum.io/downloads)distribution of Python, which already has the above packages and more included. Make sure that you select the Python 2.7 installer and not the Python 3.x installer.

**Starting the Project**

For this assignment, you can find the .zip archive containing the necessary project files as a downloadable in the **Resources** section. You may also visit our [**Projects GitHub**](https://github.com/udacity/machine-learning) to have access to all of the projects available for this Nanodegree.

The archive contains three files:

* Titanic\_Survival\_Exploration.ipynb: This is the main file where you will be performing your work on the project.
* titanic\_data.csv: The project dataset. You’ll load this data in the notebook.
* titanic\_visualizations.py: This Python script contains helper functions that will visualize the data and survival outcomes.

To open up the iPython notebook, you will need to open the Command Prompt or PowerShell (Windows), or Terminal application (Mac, Linux). Use the cd command to navigate through the file structure of your computer to where you extracted the project files. For example, on Windows you might start with cd C:\Users\username\Documents\ (substituting in the username) to get to the Documents folder. On Mac you might start with cd ~/Documents/. You can use the dir(Windows) or ls (Mac, Linux) command to list files and folders in your current directory; cd .. will move you up a directory if you get lost.

Once you’ve navigated to the folder containing the project files, you can use the commandipython notebook Titanic\_Survival\_Exploration.ipynb to open up a browser window or tab to work with your notebook. There are five questions in the notebook that you need to complete for the project.

**Project Deliverables**

You should submit the following files as your submission, packaged in a single .zip archive:

* Completed Titanic\_Survival\_Exploration.ipynb notebook with answers on all questions and cells run successfully with output.
* HTML export of the notebook. Instructions for export are at the bottom of the notebook; you may need to install the [**mistune**](https://pypi.python.org/pypi/mistune) package first, e.g. via pip install mistune in the terminal.

**Linux users**: apt-get install ipython may install an older version (1.2.1) which is not recent enough for this assignment. Instead, you may run sudo pip install jupyter followed byjupyter notebook Titanic\_Survival\_Exploration.ipynb to launch the notebook. See [**this discussion post**](https://discussions.udacity.com/t/ipython-software-installation-for-linux/163320).