Exercise 101

Business Data Analytics and Prediction (Foundations and Advanced) 2264

A. Write a data codebook

Write a data codebook to your project.

See example for a codebook, and for Python notebook that can help you understand the data in: https://github.com/shamagilli/IDC-BDA/tree/master/lesson0101_introduction

Your team

1. What is your team number? Name? _____
2. Check who are your team members, and their contact information
3. What is your team members background and strengths for the data science capabilities

The data

4. What is the link to your team's data set? _____
5. Indicate source credentials / data owner
6. Specify data copy rights (if any) and/or publication limitations

Business questions

7. Write 4 business questions that you can answer with this data.
8. Who (hypothetically) needs to review your business questions before you analyze?
9. Find an academic article(s) (patent or blog) that relates to data similar to yours.
    a. Provide full reference and link.
    b. What was the main conclusion on it?
    c. Can you look for similar conclusions in your data?

The variables (columns)

10. Per each variable describe:
    a. Variable label
    b. Variable full name or description (e.g., for surveys put the question text)
    c. Possible values and value
    d. Summary statistics
    e. Missing values

B. Setup your computer

11. Create a folder named exercises for the course exercises.
12. Check that you have **pip v. 10** on your desktop:
    a. Open a command line (in Windows, on the run line type: cmd)
    b. Click in the command line: pip --version

```
C:\Users\gillis>pip --version
pip 9.0.1 from C:\Users\gillis\AppData\Local\Continuum\Anaconda3\lib\site-packages (python 3.6)
```

c. If you need to upgrade, then type the command: pip install --upgrade pip, and then re check the version
…

```
C:\Users\gillis>pip install --upgrade pip
Collecting pip
  Downloading https://files.pythonhosted.org/packages/0f/74/ecd13431bcc456ed390b44c8a6e917c1820365cbebcb6a8974d
    100% |████████████████████████████████| 1.3MB 778kB/s
Installing collected packages: pip
  Found existing installation: pip 9.0.1
    Uninstalling pip-9.0.1:
      Successfully uninstalled pip-9.0.1
Exception:
C:\Users\gillis>pip --version
pip 10.0.1 from c:\users\gillis\appdata\local\continuum\anaconda3\lib\site-packages\pip (python 3.6)
```

13. Install Anaconda, if you do not have it yet.
    a. If you don't have Anaconda, you can install it from: https://www.anaconda.com/download/ for Python 3.6 version
    b. If you have anaconda, check that it is for Python 3.6. The checkup can be done in a command line:

```
C:\Users\gillis>python --version
Python 3.6.0 :: Anaconda custom (64-bit)
```

    c. If you have anaconda, you may upgrade it with the command: conda upgrade conda. Missing packages will be added.

```
C:\Users\gillis>conda upgrade conda
Solving environment: done

# All requested packages already installed.
```

14. Using Jupyter notebook, create a Python file (or notebook file?) named exercise101 in your exercises folder. This file will have only one Python command to print your name.

C. Set up the group Github

What is github

Github is used for code source sharing: both open source social sharing, as well as Joint group coding.

15. See the movie on what is Github in: https://www.youtube.com/watch?v=w3jLJU7DT5E
16. Go to Github.com and search all github to open source solutions related to the business questions you are addressing
    a. What was your search term? _____
    b. How many open source repository you found? _____
    c. How many of the repositories are in R? _____
    d. How many of the repositories are in Python (and Jupyter notebook)? _____

Github online

To work with github you'll need a github account (like your Facebook or LinkedIn account) and to install locally Git Bash to manage the connectivity of your desktop to github online.

17. Create your github account (if you don't have one yet)
    a. Go to **www.github.com** and **login**, or create your profile if you don't have one yet.
    b. What is your user name for your git account? _____
    c. What is the email for your git account? _____
    d. Update your user profile (e.g., add your photo)

18. Create a github repository named '**IDC BDA Exercises**'
    - Click **New Repository**. (Green button on the right)
    - Name your repository '**IDC BDA Exercises**'.
    - Write a short description.
    - Select initialize this repository with a README.
    - Click **Create Repository**.
19. Edit your new repository README file
    a. Click on the '**IDC BDA Exercises**' repository
    b. Click on **README.md** (a text will appear below the files list)
    c. Click on the **Edit** button (pencil icon)
    d. Make a change in the text
    e. Click **Commit changes**
    f. Go **Back** (or click on 'IDC BDA Exercises' in the top path)
20. Format the text
    a. Edit **README** file
    b. Format the text using formatting characters:
    c. ========= makes the line above a title format with a line beneath
    d. # on beginning of line is heading 1
    e. ## on beginning of line is heading 2
    f. See more formatting tips in: https://help.github.com/articles/basic-writing-and-formatting-syntax/#styling-text
21. Add the python notebook file exercise101.ipynb to your repository
    a. Click **Upload Files** (top button)
    b. Drag the file to upload
    c. Click **Commit changes**
    d. Go **Back** (or click on 'IDC BDA Exercises' in the top path)
22. Select a repository for the group exercises
    a. Create one repository for your group
        i. Share the repository link with all team members, and save it on your favorite links
        ii. What is the repository link? _____
    b. Edit the **README** for this repository to include:
        i. The team number and name
        ii. All group members names
    c. Upload your data file(s) to this repository

Setup your personal desktop to communicate with the group's github – one time actions

Each group member will have a copy of your project's repository on his desktop.

23. Install Git Bash from https://git-scm.com/downloads
    - If you had already git bash install, then update to latest version
24. Go to your local folder exercises, right-click and select Git-Bash Here
    - If you have a command window opened you are all set.
25. Type the command: git --version
    - Which git version are you using? _____
26. Configure your user name and email, with the commands:
    - **git config --global user.name "type_here_your_user_name"**
    - **git config --global user.email "type_here_your_email"**
27. Clone the group project repository
    a. Open your repository online
    b. Select **Clone or download** 

c. Select **Copy to repo** 
d. Go to your computer to the folder to which you would like to clone this repository. Right click and select **Git bash here**
e. Check that the prompt on the git bash is on the correct folder (you can change folders with the cd command)
f. Type the command **git clone** and then paste the address
g. Check that the repo appears as a folder where you cloned it

Ongoing work on your project

The setup commands are done once. In ongoing you wish to get code from others, and to contribute your changes.

28. On the git bash command line check that you are on the project's directory
29. Each time you'll continue working on a project perform git pull to make sure you work on the latest version
30. Run the command git pull
31. Make your changes on your folder
    - For example, edit the README file with notepad, and save the changes into the file
32. Contribute your changes in the following order
    - **git status** #See all your changes and others changes
    - **git add --all** #or git add .
    - **git commit -m** "My message" #Make meaningful description
    - **git push** #Youll need to sign with your user name and password

D. Submit exercise 101

To submit this exercise, submit in moodle only one for the team.

Submit only the link to the team github repository.

Your exercise in done if:

- All team members names appear on the github README
- The data file(s) is in this repository
- The data codebook file is in this repository
- There is one file only of ipynb in this repository, named exercise101.
- The exercise101 file, was edited by each of the team members, each changing the file print command.