Classified_Courses

April 6, 2016

In [26]:
```
import numpy as np
from datascience import *
```

In [27]:
```
# Loads the csv data file for further data cleaning
file_name = "courses.csv"
un_cleaned = Table.read_table(file_name)
courses = un_cleaned.take(np.arange(152)).drop([9,10])
courses
```

Out[27]:

Type of Course	ID	Department	Course #	Course Name
Method	1	African American Studies (AFRICAM)	101	Research Methods i
Method	2	American Studies (AMERSTD)	C134	Information Techno
Core	3	Anthropology (ANTHRO)	169A	Data Analysis and
Method	4	Anthropology (ANTHRO)	135	Paleoethnobotany:
Method	5	Anthropology (ANTHRO)	169B	Research Theory ar
Method	6	Anthropology (ANTHRO)	169C	Research Theory ar
Method	7	Architecture (ARCH)	154	Design and Comput
Method	8	Asian American Studies Program (ASAMST)	165	Research Methodol
Core	9	Bioengineering (BIO ENG)	143	Computational Met
Method	10	Bioengineering (BIO ENG)	131	Introduction to C

... (142 rows omitted)

In [108]:
```
applied = ["Research", "Analy", "Method", "Access", "Computational", "Language Processing",
           "Artificial Intelligence","Population Science", "Econo", "Optimization",
           "Geospatial", "Count", "Quality Control", "Statistics", "Behavior and Data",
           "Empiricism", "Literature and Data", "Experiment", "Time Series", "Sampling",
           "Statistical Models", "Mathematical Statistics", "Matrices and Graphs"]
meta = ["Privacy", "Ethic", "Law", "Philosophy", "Society", "Cultur", "Cities", "Mind",
        "Government", "Demography", "Policy", "Simulation", "Poverty", "Global", "Agriculture
        "Historical", "Environmental", "History", "Media", "Decision Making", "Social"]
tool = ["Program", "Query", "Database", "Programming", "Systems", "Digital", "Software",
        "Algorithm", "Processing", "Machine Learning", "Data Science", "Data Structures",
        "Unix", "Machine Structures", "Computing", "Probability", "Introduction to Statistics
is_applied = []
is_meta = []
is_tool = []
def is_category(course, terms):
    for term in terms:
        if term in course:
            return True
    return False

def array_maker(terms, added_array):
    for i in courses.column("Course Name"):
```

```
            added_array.append(is_category(i, terms))
        return added_array

    courses = courses.with_columns([
            "Applied", array_maker(applied, is_applied),
            "Meta", array_maker(meta, is_meta),
            "Tool", array_maker(tool, is_tool)
        ])
    courses
```

Out[108]:

Type of Course	ID	Department	Course #	Course Name
Method	1	African American Studies (AFRICAM)	101	Research Methods
Method	2	American Studies (AMERSTD)	C134	Information Techn
Core	3	Anthropology (ANTHRO)	169A	Data Analysis and
Method	4	Anthropology (ANTHRO)	135	Paleoethnobotany
Method	5	Anthropology (ANTHRO)	169B	Research Theory a
Method	6	Anthropology (ANTHRO)	169C	Research Theory a
Method	7	Architecture (ARCH)	154	Design and Comput
Method	8	Asian American Studies Program (ASAMST)	165	Research Methodol
Core	9	Bioengineering (BIO ENG)	143	Computational Met
Method	10	Bioengineering (BIO ENG)	131	Introduction to C

... (142 rows omitted)

In [109]:
```
courses.to_df().to_csv('classified_courses.csv', index = False)
```

In [ ]: