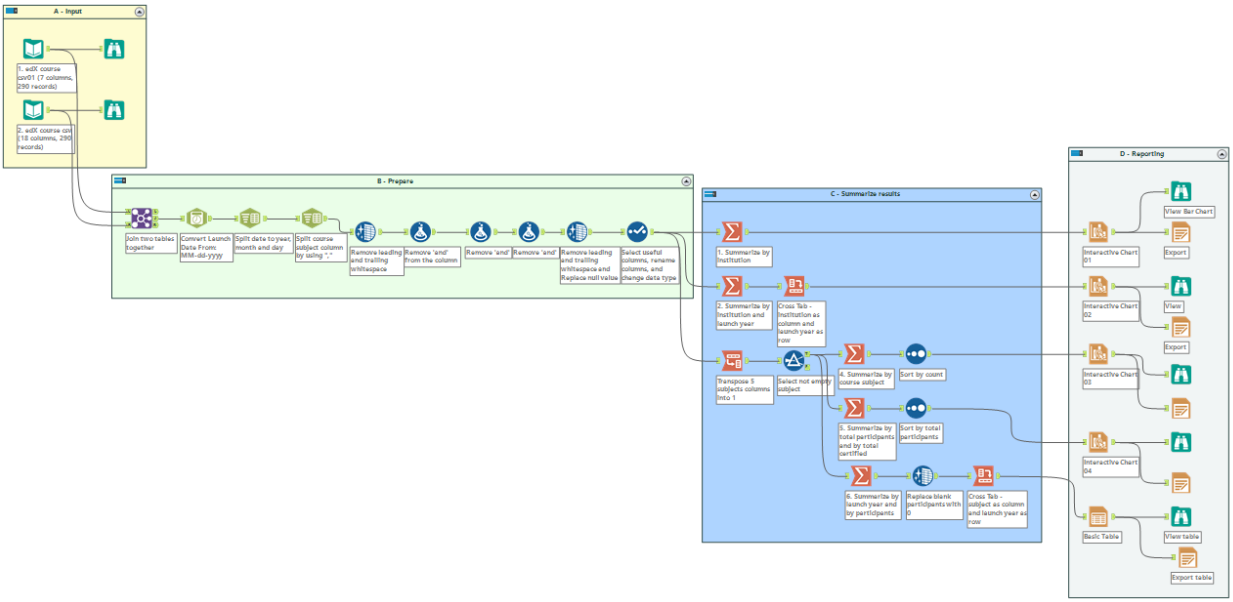
Preparing and Visualizing Data with Alteryx

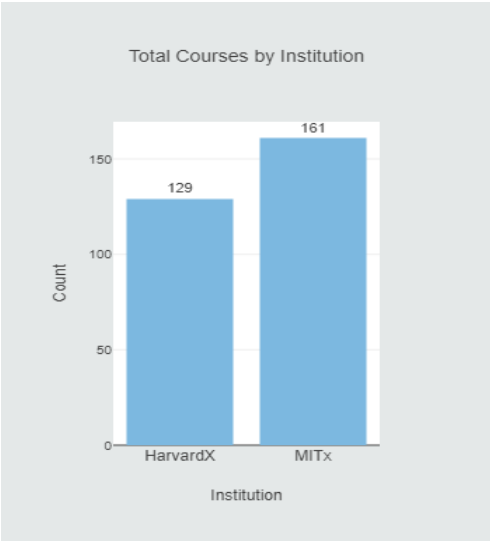
by Qi Sun

In 2012, the Massachusetts Institute of Technology (MIT) and Harvard University launched open online courses on edX, a non-profit learning platform co-founded by the two institutions. These two datasets provide data on 290 Harvard and MIT online courses, 250 thousand certifications, 4.5 million participants, and 28 million participant hours on the edX platform since 2012. The purpose of this project is to conduct a descriptive analysis on the course information by institutions, by launch years and by course subjects.

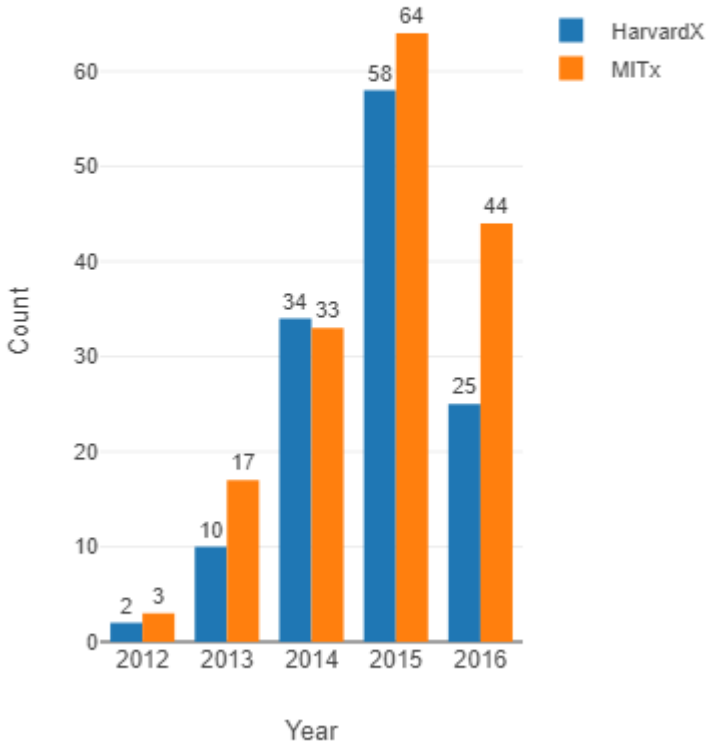
Here is a screenshot of the workflow:



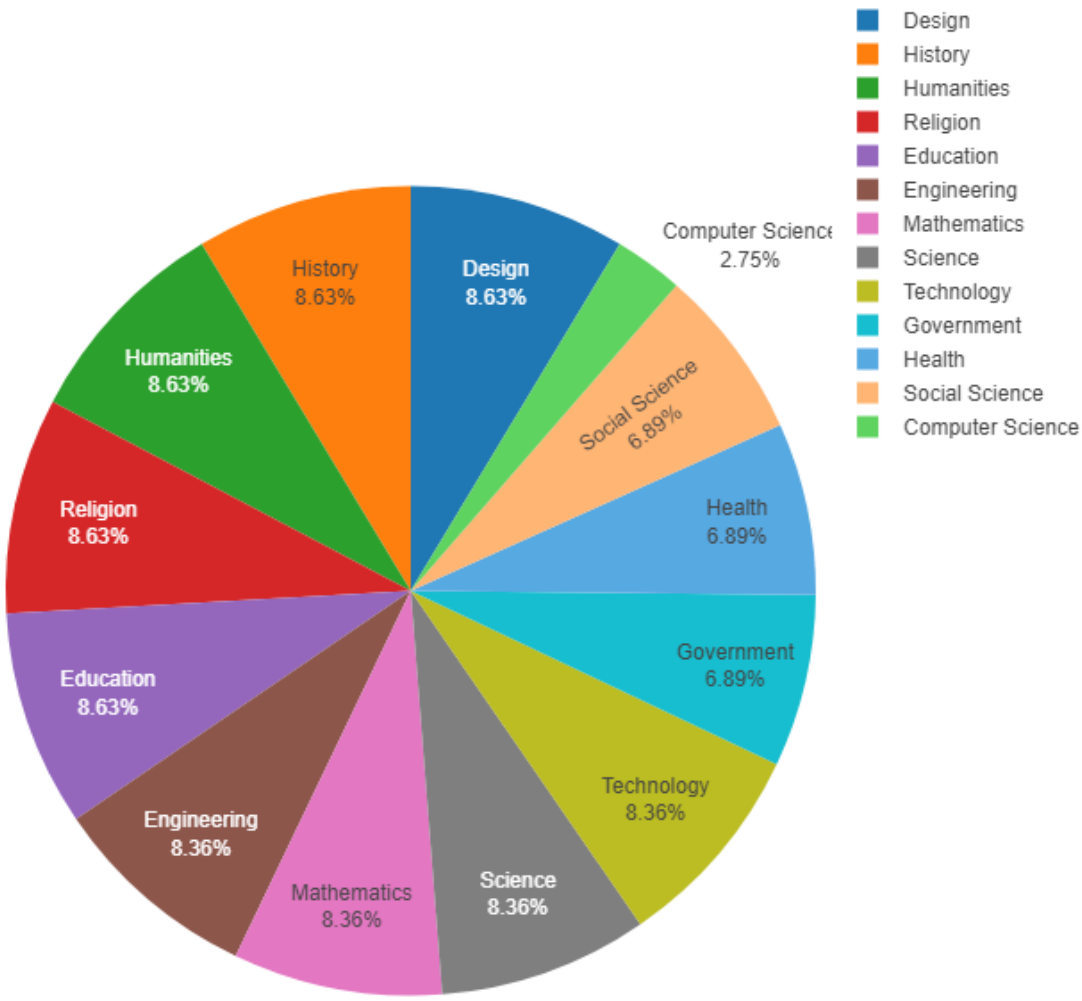
1. **The total number of courses by institution:**
   As shown below, there are 129 courses run by HarvardX and 161 courses run by MITx.



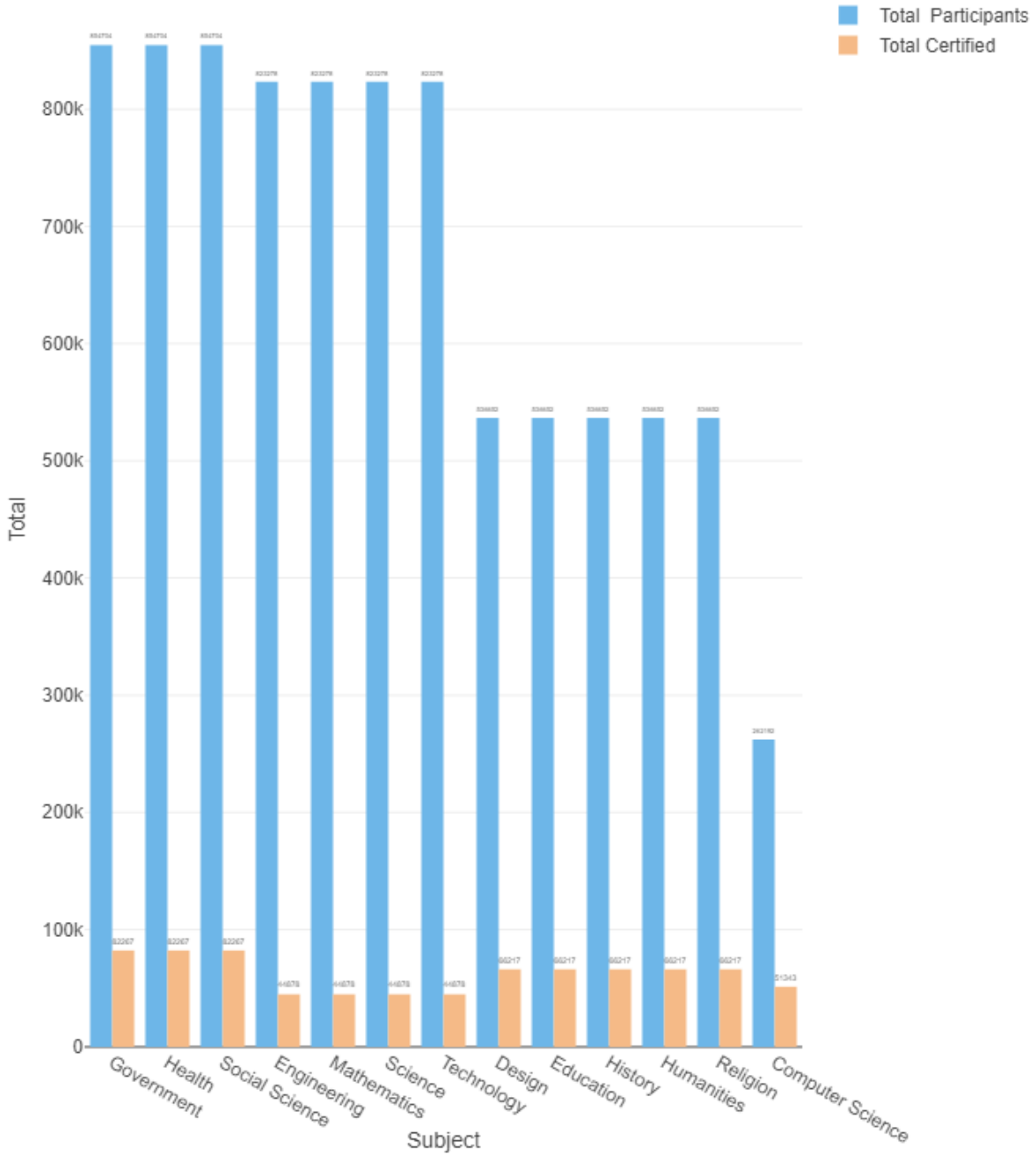
2. **The total number of courses by institution and by year:**



3. **The total number of courses by course subject:**


Design
History
Humanities
Religion
Education
Engineering
Mathematics
Science
Technology
Government
Health
Social Science
Computer Science
Design 8.63%
History 8.63%
Humanities 8.63%
Religion 8.63%
Education 8.63%
Engineering 8.36%
Mathematics 8.36%
Science 8.36%
Technology 8.36%
Government 6.89%
Health 6.89%
Social Science 6.89%
Computer Science 2.75%

4. **The total number of participants and Certified students by subject**



5. **The total number of participants and Certified students by subject and by year**

Yea	Computer Science	Desig	Educatio	Engineerin	Governme	Healt	Histor	Humanitie	Mathemati c	Religio	Scienc	Social Scienc	Technolo g
2012	0			16,663	0	0			16,663		16,663	0	16,663
2013	0	67,000	67,000	154,122	132,997	132,99 7	67,000	67,000	154,122	67,000	154,122	132,997	154,122
2014	114,952	193,05 4	193,054	218,043	175,404	175,40 4	193,05 4	193,054	218,043	193,054	218,043	175,404	218,043
2015	85,834	171,73 3	171,733	310,387	405,297	405,29 7	171,73 3	171,733	310,387	171,733	310,387	405,297	310,387
2016	61,406	104,86 5	104,865	124,063	141,006	141,00 6	104,86 5	104,865	124,063	104,865	124,063	141,006	124,063