

Research Questions:

1. What was the situation of students who discontinued their studies over the past ten years? Any differences between the level of study and school?
2. How was the employment situation of Full-time Graduates over the past decade among universities? Were there any significant changes in terms of employment preference? What were the most popular industries our graduates usually work with?
3. The Innovation and Technology Bureau (“ITB”) was formally established on 20 Nov 2015 in order to boost the Innovation and Technology industry. Were there any positive changes for that industry in terms of graduate employment preference after the setting up of ITB?

Datasets preprocessing procedure:

We have a total of 10 datasets collected from <https://cdcf.ugc.edu.hk/cdcf/indepthAnalysis.action>, based on different topics. Data sets are as follow:

- [1]"Admission_Qualification_of_First-year_Ug_Intakes_by_Age_(hc).xls" &
- [2]"Admission_Qualification_of_First-year_Ug_Intakes_by_Institution_(hc).xls" &
- [3]"Employment_Situation_of_Full-time_Graduates.xls" & [4]"First-year_Student_Intakes_(hc).xls" &
- [5]"Graduates_(hc).xls" & [6]"Movement_of_Academic_Staff_(fte).xls" &
- [7]"Staff_Number_in_Academic_Departments_(hc).xls" & [8]"Student_Enrolment_by_Age_(hc).xls" &
- [9]"Student_Enrolment_by_Institution_(hc).xls" &
- [10]"Students_Who_Discontinued_Their_Studies_(excluding_RPg)_(hc).xls"

The observation with “Research Postgraduate” would be filtered out. Depending on some dataset, we dropped the year of 2009 in analyzing the question of the discontinuation of undergraduate study to match the “10 years” length we designed, meanwhile the others are using the longevity of a range between 2009 to 2019.

Note: University abbreviation list:

The Chinese University of Hong Kong (“CUHK”); The University of Hong Kong (“HKU”); The Hong Kong Polytechnic University (“PolyU”); City University of Hong Kong (“CityU”); The Hong Kong University of Science and Technology (“HKUST”); Hong Kong Baptist University (“HKBU”); The Education University of Hong Kong (“EdUHK”); Lingnan University (“LU”)

Methodology:

In this project, we used RStudio to visualize data and figure out answers to the research questions above.



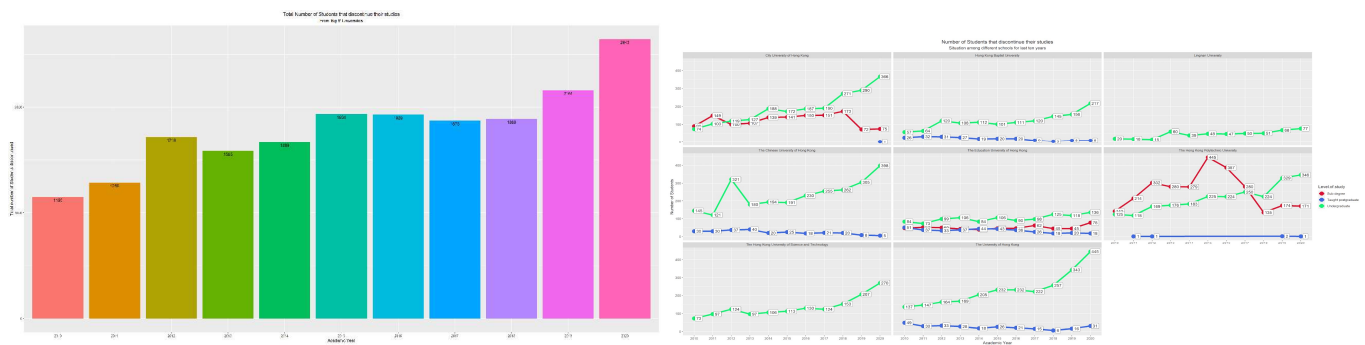
As shown in the flowchart, we draw conclusions for each question by going through these 5 stages:

1. Import Data – Read in datasets, observe data characteristics, and check inconsistencies/abnormalities
2. Data Cleaning
 - a. Eliminate abnormal/null values (e.g. @ in columns which expect numeric values)
 - b. Re-categorize observations (e.g. classify occupations based on different industries instead)
 - c. Convert data type of specific columns for transformation in the next stage.

- d. Group by data, filter, mutate columns, aggregate based on grouping
3. Data Transformation
 - a. Select relevant observations for visualization
 - b. Use Groupby and Aggregate command to summarize observations
4. Visualization
 - a. Use the ggplot2 package to gain many insights into data.
 - b. Draw line charts, pie charts, and bar charts, lollipop
 - c. assign colors as we have some graphs containing over 20 colors. Using library RColorBrewer and Polychrom, scale_fill_brewer to call the color out.
5. Explanation – Describe trend/pattern of plots and explain the phenomenon behind them.

Results:

Number of Students that Discontinued Their Studies, by Year and by University



The graph on the left illustrated the total number of discontinuities by year. The figures are relatively stable from 2015 to 2018. However, there was a jump in 2019 and 2020.

According to the graph on the right, in 2020, HKU had the largest number of discontinuity of undergraduates (445) while CUHK was in the second largest position (398). CityU has 366 which took the 3rd largest place. The percentage of subtotal numbers of these three universities in 2020 is 44.6% of the total observations.

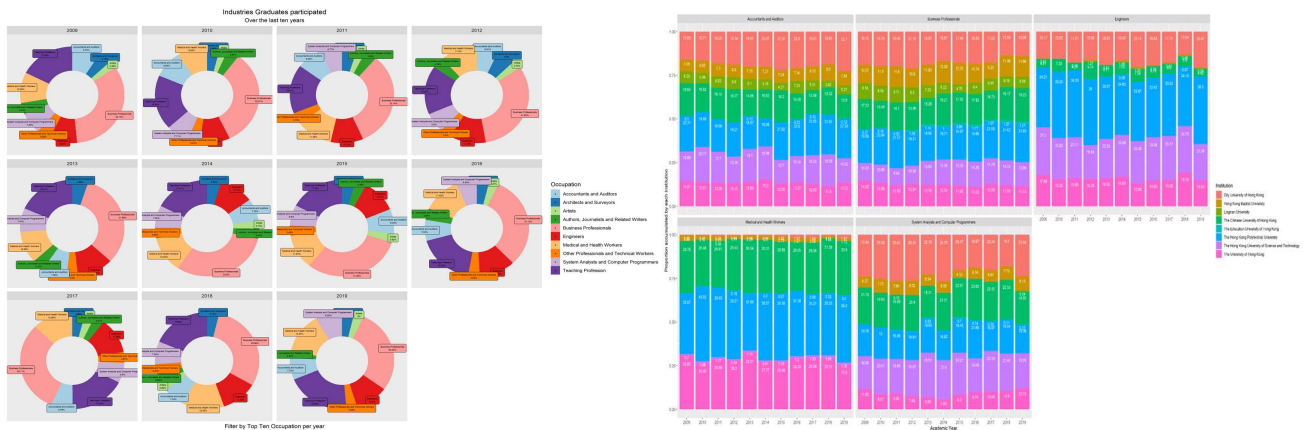
Job Categories Graduates Participated



These three graphs showed the full-time jobs distribution of graduates who are employed once after graduation in the past decade. Business professionals have many features that attract graduates, with more than 30,000 (25%) of the total samples who are employed in this field. Almost 20,000 (9.5%) of them cite engineers as another major employment attraction and chose to work in this area.

The graph in the middle and right indicated the biggest proportion of graduates worked in the business sector for most of the universities except for those graduates from EduHK.

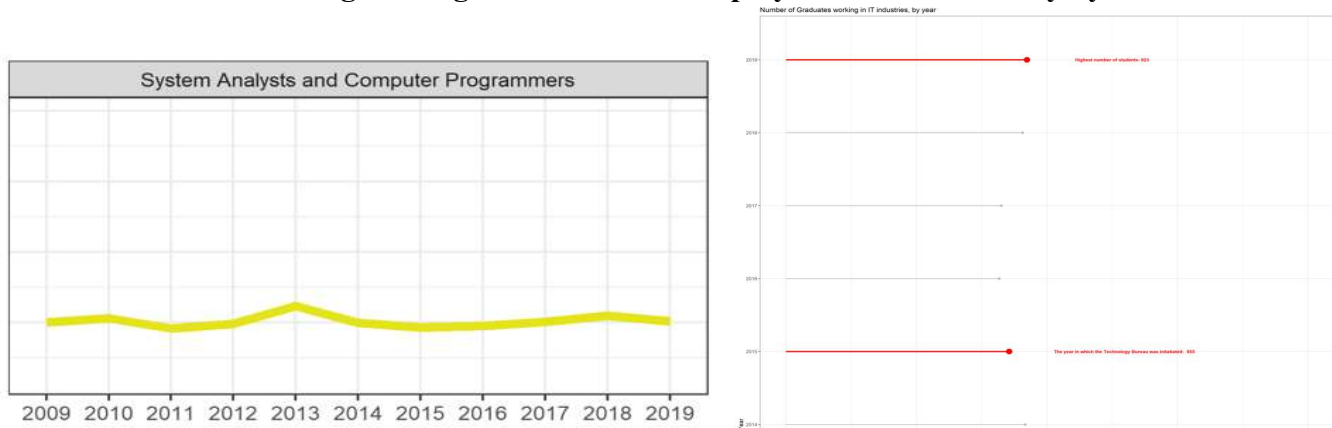
Number of Graduates from UGC-Funded Institution, by Year and by top-5 industries



The left graph indicated the most popular industries the graduates usually work with yearly. It is shown that business professional is the most competitive career in every single year. It provided an offer to one-fourth of the whole university graduates yearly. Conversely, there were only around 6% of graduates who were going to the system and computer programming field yearly.

The right graph features the top 5 industries most graduates work in and indicates which university these graduates came from. In terms of business professionals, most of the graduates are from PolyU (over 21%), CUHK (over 19%), CityU (over 13%), and HKUST (over 13%).

Percentage Change of Number of Employment in IT Industry by Year



This graph illustrates that there was no significant change in terms of the number of graduates employed by IT companies before and after 2015, in which the ITB was formally established. There was no significant change in terms of the number of graduates employed by IT companies before and after 2015, in which the ITB was formally established. As a result, it seems there are not any easy-capture impacts on the IT industry after the establishment of ITB.

Reference Links:

<https://cdcf.ugc.edu.hk/cdcf/indepthAnalysis.action>
<https://r-graphics.org/recipe-line-graph-points>
<https://r-charts.com/part-whole/stacked-bar-chart-ggplot2/>
<https://www.r-graph-gallery.com/48-grouped-barplot-with-ggplot2.html>
<https://www.datasciencemadesimple.com/remove-duplicate-rows-r-using-dplyr-distinct-function/>