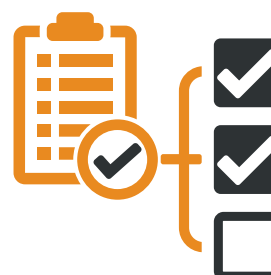**OCTOBER, 2021**

# Data Analysis Competition Submission

Economics Association,

The American University in Cairo

**SUBMITTED BY:** **Youssef M. Hussien**

**SID: 900183162**

youssefhussien@aucegypt.edu

# EAComptetion_Submission_YoussefHussien_900183162

October 31, 2021

Welcome to the Economic Association Competition Data Analysis! This project is part of the Economic Association Competition offered through The American University in Cairo and Solved by:

**Youssef M. Hussien**

SID: 900183162

Email: youssefhussien@aucegypt.edu

---

# 1 Competition Submission: Investigate the "ELMPS-2012" Dataset

## 1.1 Table of Contents

# 2 Introduction

## 2.1 Dataset Description

The dataset we are analysing throughout this report is the "ELMPS-2012" dataset that stands for the Egyptian Labor Market Panel Survey in 2012. The data provides information about a representative sample of individuals and households in Egypt in 2012. Therefore, we can consider the facts that is presented as a true representation of the Egyptian Market in reality.

**List of attributes "Characteristics of the dataset":**

1. **indid**: Individual ID

2. **hhid**: Household ID

3. **pn**:

4. **gov**: Governorate of Location

5. **region**: Region of location of Individual

6. **urban**: Urban OR Rural

7. **sex**

8. **brthyr**: Birth Year of individual

9. **age**

10. **marital**: Marital Status of individual

11. **educ**: Education level of individual

12. **fteducst**: Father Education of individual

13. **crecac1d**: Sector of Activity of individual

14. **ttmonwg3**: Total Monthly Wage detected for last three months

15. **q6101_07**: Job Stability of individual

## 2.2 Research Question(s) for Analysis

1. What is ELMPS Data? How is it collected?

2. Present All Facts (with bullet points) that can be exported from the data.

3. Present a table that covers information about each governorate and what is markable for each one. In other words:

   - The First column in the table is the name of the governorate.
   - The second column in the table is its rank in terms of illiterate, higher education,....

4. Present Facts about each governorate in terms of Marital Status as a share of the population. (Population data can be exported from the CAPMAS website, accessible at the library).

5. Any relationship between Father Education and Individual Education, per governorate in Egypt? Search about the economic theory that can support your findings and present it.

# 3 Data Wrangling (Looking at the Data)

In this section of the report, we will load in the data, check for cleanliness, and then trim and clean the dataset for analysis.

## 3.1 Get the data and explore some of its main properities

Explore the overall structure of the dataset

```
[ ]: Image("drive/My Drive/Economics Data Analysis Competition/Images/overall.png")
```

[ ]:

| | A | B | C | D | E | F | G | H | I | J | K | L | M | N | O |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | indid | hhid | pn | gov | region | urban | sex | brthyr | age | marital | educ | fteducst | crecac1d | ttmonwg3 | q6101_07 |
| 2 | 12010001101 | 120100011 | 1 | Cairo | Gr. Cairo | urban | female | 1963 | 49 | divorced | Intermediate | Less than Intermediate | | | permanent |
| 3 | 12010001102 | 120100011 | 2 | Cairo | Gr. Cairo | urban | female | 1929 | 83 | widowed(er) | Illiterate | Reads & Writes | | | |
| 4 | 12010002102 | 120100021 | 2 | Cairo | Gr. Cairo | urban | male | 1988 | 24 | never married | Intermediate | Illiterate | F:Construction | | casual |
| 5 | 12010002103 | 120100021 | 3 | Cairo | Gr. Cairo | urban | male | 1986 | 26 | married | Illiterate | Illiterate | F:Construction | | casual |
| 6 | 12010002101 | 120100021 | 1 | Cairo | Gr. Cairo | urban | female | 1949 | 63 | widowed(er) | Illiterate | Illiterate | | | |
| 7 | 12010002104 | 120100021 | 4 | Cairo | Gr. Cairo | urban | female | 1989 | 23 | married | Less than Interme | Illiterate | | | |
| 8 | 12010003103 | 120100031 | 3 | Cairo | Gr. Cairo | urban | male | 1995 | 16 | less than minimum | Less than Interme | Reads & Writes | | | |
| 9 | 12010003104 | 120100031 | 4 | Cairo | Gr. Cairo | urban | male | 1998 | 14 | less than minimum | Less than Interme | Reads & Writes | | | |
| 10 | 12010003101 | 120100031 | 1 | Cairo | Gr. Cairo | urban | male | 1970 | 42 | married | Reads & Writes | Illiterate | F:Construction | | 2328 temporary |

## 3.2 Print the names of the columns of the dataset and their properities

```
[ ]: print("These are the columns of the dataset and some small insights about them:
      ↪\n")
      Image("drive/My Drive/Economics Data Analysis Competition/Images/columns.png")
```

These are the columns of the dataset and some small insights about them:

[ ]:

```
RangeIndex: 49186 entries, 0 to 49185
Data columns (total 15 columns):
 #   Column     Non-Null Count   Dtype
---  ------     --------------   -----
 0   indid      49186 non-null   int64
 1   hhid       49186 non-null   int64
 2   pn         49186 non-null   int64
 3   gov        49186 non-null   object
 4   region     49186 non-null   object
 5   urban      49186 non-null   object
 6   sex        49186 non-null   object
 7   brthyr     49184 non-null   float64
 8   age        49186 non-null   int64
 9   marital    49186 non-null   object
 10  educ       36787 non-null   object
 11  fteducst   40666 non-null   object
 12  crecac1d   16307 non-null   object
 13  ttmonwg3   7644 non-null    float64
 14  q6101_07   13628 non-null   object
dtypes: float64(2), int64(4), object(9)
```

---

**So from the above data it seams that:**

1. We don't have empty entries for the following columns:

- indid

- hhid

- pn

- gov

- region

- urban

- sex

- marital

2. The following columns or characteristics have a large amount of missing data:

- educ: 12399 missing entries

- fteducst: 8520 missing entries

- crecac1d: 32879 missing entries

- ttmonwg3: 32879 missing entries

- ttmonwg3: 41542 missing entries

- q6101_07: 35558 missing entries

3. There are only two individuals whose birth year is missing

**Please note that** these data points are missing, but for instance this doesn't mean that if educ data point is missing that this individual is illetrate. No, it means that its education entry is missing only.

Now let's investigate some numerical properies of each column of the dataset

```
[ ]: Image("drive/My Drive/Economics Data Analysis Competition/Images/num_props.png")
```

```
[ ]:
```

Investigating some numerical properities of the dataset (numerical columns) :

| | indid | hhid | pn | brthyr | age | ttmonwg3 |
|---|---|---|---|---|---|---|
| count | 4.918600e+04 | 4.918600e+04 | 49186.000000 | 49184.000000 | 49186.000000 | 7644.000000 |
| mean | 1.217205e+10 | 1.217205e+08 | 2.989489 | 1985.041802 | 26.308258 | 3603.677917 |
| std | 8.155263e+07 | 8.155263e+05 | 1.897311 | 19.908915 | 19.932386 | 4244.501672 |
| min | 1.201000e+10 | 1.201000e+08 | 1.000000 | 1906.000000 | 0.000000 | 0.000000 |
| 25% | 1.213023e+10 | 1.213023e+08 | 2.000000 | 1973.000000 | 9.000000 | 1800.000000 |
| 50% | 1.218036e+10 | 1.218036e+08 | 3.000000 | 1988.000000 | 24.000000 | 2700.000000 |
| 75% | 1.224046e+10 | 1.224046e+08 | 4.000000 | 2002.000000 | 38.000000 | 4050.000000 |
| max | 1.229010e+10 | 1.229010e+08 | 21.000000 | 2012.000000 | 106.000000 | 105000.000000 |

From the above numerical results we can note a couple of things:

1. We have a high variation in the birth years, as it has a high standard deviation. However, disregarding the standard deviation we can find simply that the minimum and maximum of birth years gives the following two facts:

- The oldest individual we have in our data set is born at 1906

- The youngest individual we have in our data set is born at 2012

2. The maximum age we have is 106 years old, which makes sense as the oldest is born at 1906 and the dataset is at 2012.

3. The minimum age we have is 0 years old, which makes sense as the youngest is born at 2012 and the dataset is at 2012. But these data points are of no importance. But let's dig deeper in the next sections.

---

Now we will look at the eductaion attribute **'educ'** and see the count of individuals of each educational category or level.

```
[ ]: Image("drive/My Drive/Economics Data Analysis Competition/Images/educ_count.
     ↪png")
```

[ ]:

The number of individuals per education-category:

|  | educ |
|---|---|
| Intermediate | 10277 |
| Illiterate | 8860 |
| Less than Intermediate | 8498 |
| University | 4329 |
| Reads & Writes | 3671 |
| Above Intermediate | 929 |
| Post-Graduate | 223 |

Therefore, from the above data, we have

1. 10277 indiviudals who have *Intermediate* education.

2. 8860 indiviudals who have *Illetrate* .

3. 8498 indiviudals who have *Less Than Intermediate* education.

4. 4329 individuals who have *University* level education.

5. 3671 individuals who have *Read and Write* only.

6. 929 indiviudals who have *Above Intermediate* education.

7. 223 indiviudals who have *Above Post-graduate* level education.

8. And 12399 missing entries for education as mentioned earlier.

Now we will look at the governorate attribute 'gov' and see the count of occurences of each governorate in our dataset.

```
Image("drive/My Drive/Economics Data Analysis Competition/Images/gov_count.png")
```

[ ]:

The count of occurences of each governorate in our dataset:

|  | gov |
| --- | --- |
| Cairo | 4086 |
| Suhag | 3313 |
| Menia | 3121 |
| Asyout | 3103 |
| Sharkia | 3053 |
| Qena | 2715 |
| Gharbia | 2678 |
| Dakahlia | 2654 |
| Beni-Suef | 2561 |
| Behera | 2552 |
| Alex. | 2464 |
| Giza | 2397 |
| Kafr-Elsheikh | 2383 |
| Fayoum | 2215 |
| Kalyoubia | 2165 |
| Ismailia | 1967 |
| Aswan | 1695 |
| Menoufia | 1473 |
| Damietta | 1469 |
| Suez | 484 |
| Luxur | 376 |
| Port-Said | 262 |

## 3.3 Exploring Age Range and Age Groups

Now, we need to look at outliers and make sure that everything is correct.

The first thing we will do is to look at the age distribution we have and make sure that we don't have anything that is not reasonable, i.e having a negative age or having a very large age.

```
[ ]: Image("drive/My Drive/Economics Data Analysis Competition/Images/age_graph.png")
```

[ ]:
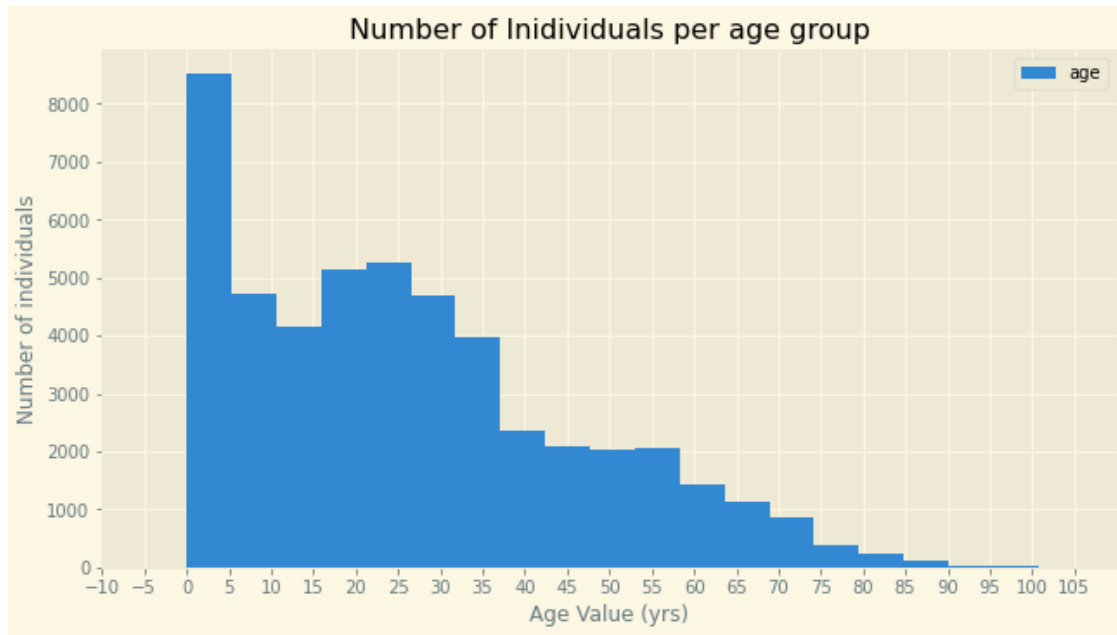


**So from the above graph we note the following:**

1. The hgihest age group represented in our data are children from less than or equal Five years old. Their total count is more than 8000 points of our data.

2. After the age of 25, or approximately or 30, the number of individuals of those groups decrease. Which means that the data was collected for young aged individuals.

---

---

## 4 Answering the Competition Questions

### 4.1 Research Question 1 ( What is ELMPS Data? How is it collected?)

#### 4.1.1 What is ELMPS Data ?

The Egypt Labor Market Panel Survey (ELMPS) is a publicly-available nationally representative longitudinal household survey. The 2012 survey is the third in a series that began with the Egypt Labor Market survey of 1998 and continued with the ELMPS 2006. It is a multipurpose household survey that inquires about the schooling, housing, marriage and employment of individuals among a number of other topics. The 2012 sample consisted of 12,060 households, including 6,752 original households. The ELMPS is a wide-ranging, nationally representative panel survey that covers topics such as parental background, education, housing, access to services, residential

mobility, migration and remittances, time use, marriage patterns and costs, fertility, women's decision making and empowerment, job dynamics, savings and borrowing behavior, the operation of household enterprises and farms, besides the usual focus on employment, unemployment and earnings in typical labor force surveys. (Assaad and Krafft, 2013)

### 4.1.2 How is it collected?

The Egypt Labor Market Panel Survey, carried out by the Economic Research Forum (ERF) in cooperation with Egypt's Central Agency for Public Mobilization and Statistics (CAPMAS) since 1998. ERF in cooperation with CAPMAS have carried out The Egypt Labor Market Panel Survey (ELMPS) during the years 1988, 1998, 2006, 2012 and 2018.

The 2012 data collection process proceeded in two phases. First, in late 2011, an enumeration phase was undertaken. This phase focused on locating households and individuals from the 2006 sample. If households or individuals had moved, every effort was made to collect current contact information. Additionally, the refresher sample was designed to over-sample high-migration areas, and refresher sample PSUs and households were randomly selected based on this sampling approach.

The 2012 ELMPS was implemented by 39 teams in the field, each consisting of one supervisor, one reviewer, and four enumerators. Additionally, there were two teams undertaking quality control. All interviewers were trained for 10 days by the technical director and CAPMAS prior to fielding the survey. The fielding of the full 2012 survey took place from March 1, 2012 to June 10, 2012, with more than 90% of households and individuals surveyed during March and April. Desk review, coding, and data entry and validation at CAPMAS occurred after fielding.

---

### 4.2 Research Question 2 ( Present All Facts (with bullet points) that can be exported from the data. )

Different facts about the data have been written throughout this notebook. I did write some facts about the data in the Data Wrangline section, as well as in this section which is full of facts about the data. I will list some of them here, but please feel free to check the full notebook to understand the different facts drawn about this data:

1. We don't have empty entries for the following columns (indid, hhid, pn, gov, region, urban, sex, marital)

2. The following columns have a large amount of missing data (educ, fteducst, crecac1d, ttmonwg3, ttmonwg3, q6101_07)

3. We have a high variation in the birth years, as it has a high standard deviation. However, disregarding the standard deviation we can find simply that the minimum and maximum of birth years gives the following two facts:

   - The oldest individual we have in our data set is born at 1906

   - The youngest individual we have in our data set is born at 2012

4. The maximum age we have is 106 years old, which makes sense as the oldest is born at 1906 and the dataset is at 2012.

5. The minimum age we have is 0 years old, which makes sense as the youngest is born at 2012 and the dataset is at 2012. But these data points are of no importance. But let's dig deeper in the next sections.

6. The hgihest age group represented in our data are children from less than or equal Five years old. Their total count is more than 8000 points of our data.

7. After the age of 25, or approximately or 30, the number of individuals of those groups decrease. Which means that the data was collected for young aged individuals.

8. Suhag and Menia had the highest illetracy percentage with 9%. But Suhag was higher only with a couple of individuals. They were followed by Asyout

9. Cairo is the highest governorate in terms of educated people on all levels of education except for Above Intermediate Education in which many other governorates exceed it.Luxor is the lowest in Intermediated Education. Port-Said is the lowest in Above Intermediate Education, the lowest in Above Intermediate Education, the lowest in Post Grauate Education and the lowest in Read and Write Education. Kalyobia is the lowest in University level Education.

10. Cairo Was the highest governorate in educated parents count, with 3602 parent. While, Port-Said again was the lowest governorate in educated parents count, with only 227 parent.

11. There are six cases or states of Marital Status can be one of the following 6 cases (according to the dataset): Contractually Married, Divorced, Less Than Miniumum Age, Married ,Never Married and Widowed(er).

12. Cairo has the highest marriage count in all governorates with 1765 marriages.

13. Divorces count is low. But count of divorce is very simliar in all governorates except for Cairo which is the highest with approximately 64 divorces which is more than triple or double all other governorates.

14. Again Cairo is the highest governorate in numbers of widows and never married individauls, with counts equal to 287 and 734 respectively.

15. Sector Of Activity of an individual can be one of 21 different cases i.e types. Thus, it will be tedious to list the results of this part a number of points. So please check *Sector Of Activity Vs Governorates* subsection under the Third Research Question Section.

16. Governorates with High Fathers Education are High in Individuals Education count. Thus, it can bee seen afterwards that Fathers education is directly proportional to Individuals education. There are a number of supporting theories about this relationship, which are stated in the Fifth Research Question section.

### 4.3 Research Question 3 (Present a table that covers information about each governorate and what is markable for each one.)

In other words:

- The First column in the table is the name of the governorate.

- The second column in the table is its rank in terms of illiterate, higher education,....)

**Instead of doing so,** I saw that it is more appropriate to compare governorates to a number of attributes, each on its own. In other words this section include the following subsections:

- Governorates Vs Illetracy Percentages

- Governorates Vs Education Levels' Percentages

- Governorates Vs Father Education

- Governorates Vs Education Total (Individuals Education and Father Education)

- Governorates Vs Marital (Parent and Individuals)

- Governorates Vs Sector of Activity

#### 4.3.1 Governorates Vs Illiteracy

The first thing I will compare is the number of illetrate individuals we have per governorate. In other words, we will copmare first the illetracy percentage in each governorate.

In the following table we compare the number of illetrate individuals per governorate.

```
Image("drive/My Drive/Economics Data Analysis Competition/Images/
 ↪illetracyGovTable.png")
```

[ ]:

| 1 | Count for Educ | Count of Education | |
|---|---|---|---|
| 2 | Governorate | Illiterate | |
| 3 | Alex. | 240 | |
| 4 | Aswan | 253 | |
| 5 | Asyout | 746 | |
| 6 | Behera | 543 | |
| 7 | Beni-Suef | 583 | |
| 8 | Cairo | 463 | |
| 9 | Dakahlia | 432 | |
| 10 | Damietta | 212 | |
| 11 | Fayoum | 442 | |
| 12 | Gharbia | 447 | |
| 13 | Giza | 413 | |
| 14 | Ismailia | 228 | |
| 15 | Kafr-Elsheikh | 472 | |
| 16 | Kalyoubia | 326 | |
| 17 | Luxur | 59 | |
| 18 | Menia | 769 | |
| 19 | Menoufia | 240 | |
| 20 | Port-Said | 17 | |
| 21 | Qena | 622 | |
| 22 | Sharkia | 546 | |
| 23 | Suez | 34 | |
| 24 | Suhag | 772 | |
| 25 | Grand Total | 8859 | |

This chart plots the previous table, but in chart-form.

```
[ ]: Image("drive/My Drive/Economics Data Analysis Competition/Images/
     ↪illetracyGovChart.png")
```
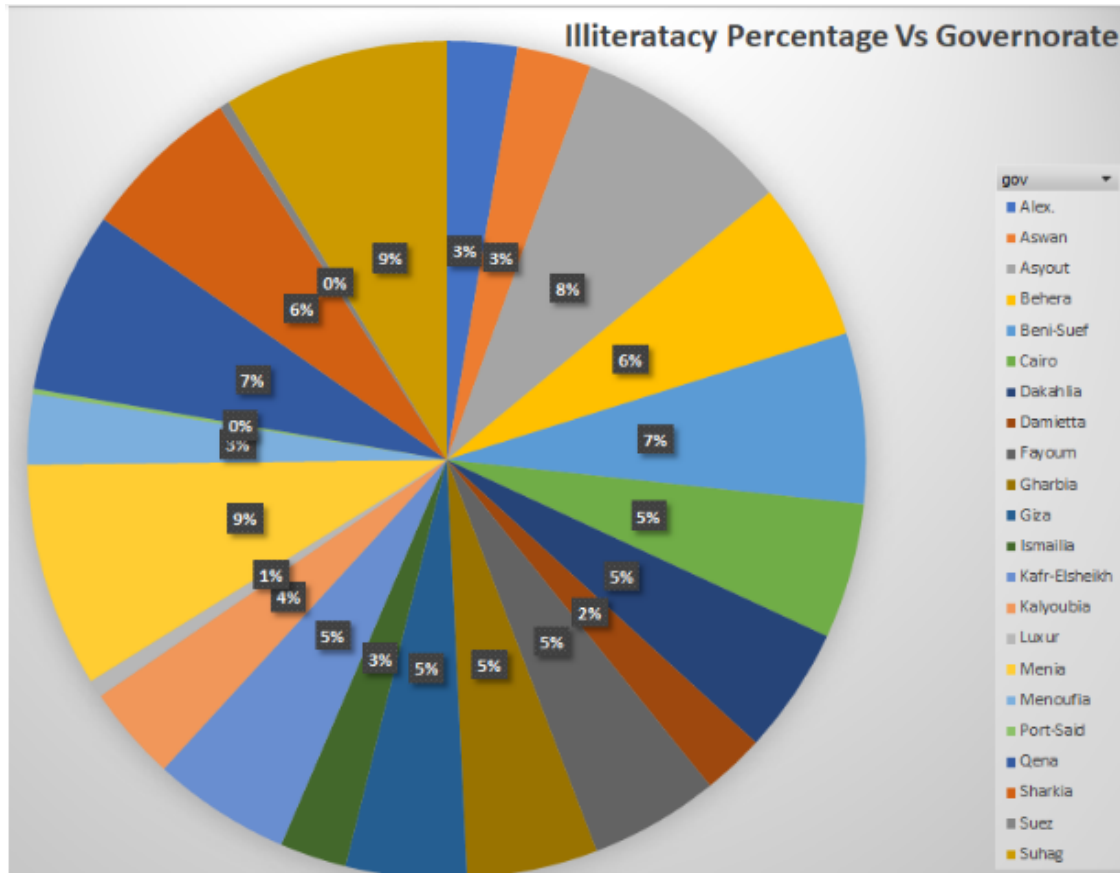
[ ]:
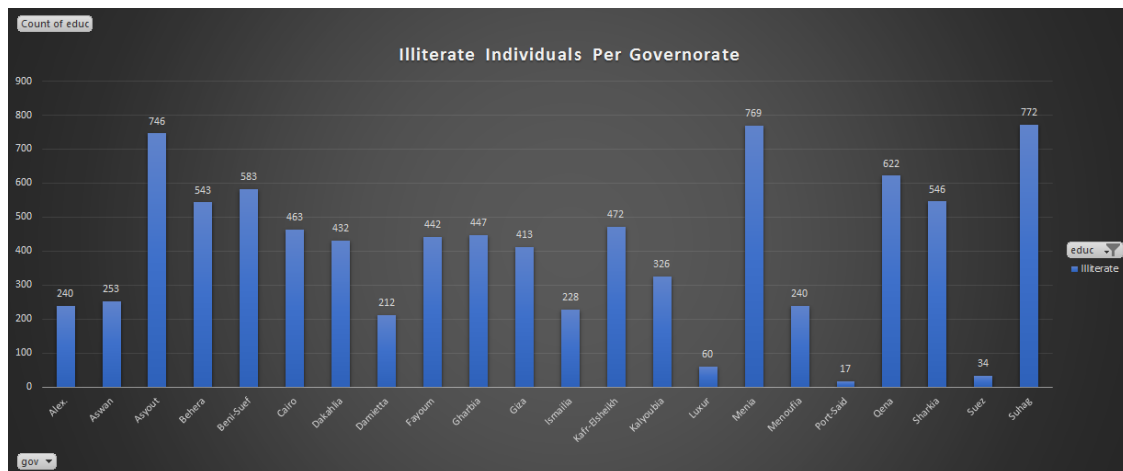


To understand this more, let's view the illetracy but in terms of percentages. The following pie chart gives us a nice representation of the illetracy percetages per governorate.

12

```
[ ]: Image("drive/My Drive/Economics Data Analysis Competition/Images/
     ↪illetracyGovPie.png")
```

[ ]:



A number of interesting facts are obvious from the above data:

1. Suhag and Menia had the highest illetracy percentage with 9%. But Suhag was higher only with a couple of individuals as seen from the chart and the table.

2. They were followed by Asyout.

### 4.3.2 Governorates Vs. Education Levels Percentages

In this section we present the different count of different education levels per governorate. In the next table and chart, each education level (other than illetracy) is mapped to its count per governorate to compare all governorates to each other.

In the following table we compare the number of individuals for each education level or category per governorate.

```
[ ]: Image("drive/My Drive/Economics Data Analysis Competition/Images/
     ↪educlevelsGovTable.png")
```

[ ]:

| | Count for Educ Count of Education | | | | | | |
|---|---|---|---|---|---|---|---|
| | Governorate | Above Intermediate | Intermediate | Less than Intermediate | Post-Graduate | Reads & Writes | University | Grand Total |
| 3 | Alex. | 89 | 508 | 592 | 13 | 181 | 325 | 1708 |
| 4 | Aswan | 33 | 443 | 286 | 2 | 130 | 144 | 1038 |
| 5 | Asyout | 39 | 557 | 508 | 4 | 226 | 161 | 1495 |
| 6 | Behera | 40 | 583 | 357 | 2 | 166 | 199 | 1347 |
| 7 | Beni-Suef | 27 | 521 | 425 | 5 | 195 | 122 | 1295 |
| 8 | Cairo | 124 | 841 | 738 | 48 | 272 | 841 | 2864 |
| 9 | Dakahlia | 56 | 611 | 471 | 16 | 213 | 212 | 1579 |
| 10 | Damietta | 30 | 315 | 269 | 3 | 124 | 123 | 864 |
| 11 | Fayoum | 33 | 438 | 359 | 9 | 186 | 119 | 1144 |
| 12 | Gharbia | 47 | 604 | 460 | 8 | 200 | 248 | 1567 |
| 13 | Giza | 51 | 436 | 472 | 18 | 161 | 225 | 1363 |
| 14 | Ismailia | 32 | 499 | 333 | 25 | 149 | 168 | 1206 |
| 15 | Kafr-Elsheikh | 37 | 555 | 394 | 9 | 183 | 165 | 1343 |
| 16 | Kalyoubia | 52 | 423 | 391 | 7 | 154 | 184 | 1211 |
| 17 | Luxur | 14 | 81 | 53 | | 23 | 41 | 212 |
| 18 | Menia | 32 | 488 | 482 | 15 | 278 | 204 | 1499 |
| 19 | Menoufia | 19 | 358 | 249 | 7 | 86 | 131 | 850 |
| 20 | Port-Said | 15 | 64 | 27 | 1 | 17 | 74 | 198 |
| 21 | Qena | 44 | 580 | 472 | | 208 | 148 | 1452 |
| 22 | Sharkia | 46 | 659 | 504 | 26 | 221 | 263 | 1719 |
| 23 | Suez | 32 | 150 | 70 | 2 | 20 | 52 | 326 |
| 24 | Suhag | 37 | 561 | 583 | 3 | 278 | 180 | 1642 |
| 25 | Grand Total | 929 | 10275 | 8495 | 223 | 3671 | 4329 | 27922 |

This chart plots the previous table, but in chart-form.

```
[ ]: Image("drive/My Drive/Economics Data Analysis Competition/Images/
     ↪educlevelsGovChart.png")
```

[ ]:
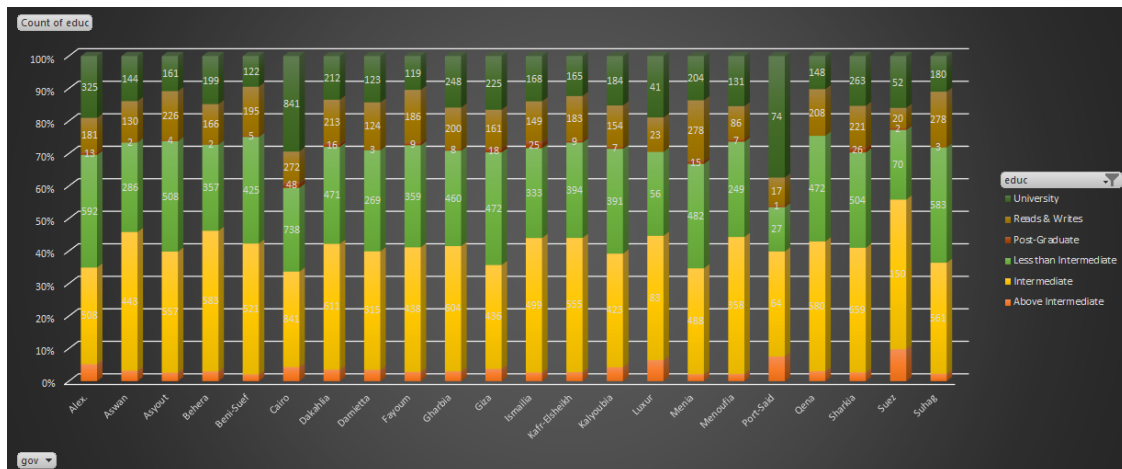


**A few FACTS arise from this chart:**

1. Cairo is the highest governorate in terms of educated people on all levels of education except for Above Intermediate Education in which many other governorates exceed it.

2. Port-Said is the lowest in Above Intermediate Education.

3. Luxor is the lowest in Intermediated Education.

4. Port-Said is the lowest in Above Intermediate Education.

5. Port-Said is the lowest in Post Grauate Education.

6. Port-Said is the lowest in Read and Write Education.

7. Kalyobia is the lowest in University level Education.

### 4.3.3 Governorate Vs. Father Education

In this third comparison, we compare and present how the count of father education differ from a governorate to the other. We will present a table showing the number of educated fathers for each governorate and then present this in a chart for ease of visualisation.

In the following table we compare the number of educated fathers per governorate.

```
[ ]: Image("drive/My Drive/Economics Data Analysis Competition/Images/
     ↪PT_ALLEducGovTable.png")
```

[ ]:

| 3 | Governorate ▾ | Parents Educated Count | Educated Individuals Count |
|---|---|---|---|
| 4 | Alex. | 2100 | 1948 |
| 5 | Aswan | 1425 | 1291 |
| 6 | Asyout | 2516 | 2241 |
| 7 | Behera | 2090 | 1890 |
| 8 | Beni-Suef | 2092 | 1878 |
| 9 | Cairo | 3602 | 3327 |
| 10 | Dakahlia | 2225 | 2011 |
| 11 | Damietta | 1187 | 1076 |
| 12 | Fayoum | 1770 | 1586 |
| 13 | Gharbia | 2209 | 2014 |
| 14 | Giza | 1940 | 1776 |
| 15 | Ismailia | 1606 | 1434 |
| 16 | Kafr-Elsheikh | 1992 | 1815 |
| 17 | Kalyoubia | 1749 | 1537 |
| 18 | Luxur | 308 | 277 |
| 19 | Menia | 2548 | 2268 |
| 20 | Menoufia | 1212 | 1090 |
| 21 | Port-Said | 227 | 215 |
| 22 | Qena | 2254 | 2074 |
| 23 | Sharkia | 2531 | 2265 |
| 24 | Suez | 404 | 360 |
| 25 | Suhag | 2679 | 2414 |
| 26 | **Grand Total** | **40666** | **36787** |

This chart plots the previous table, but in chart-form.

```
[ ]: Image("drive/My Drive/Economics Data Analysis Competition/Images/PTEducGovChart.
     ↪png")
```

[ ]:
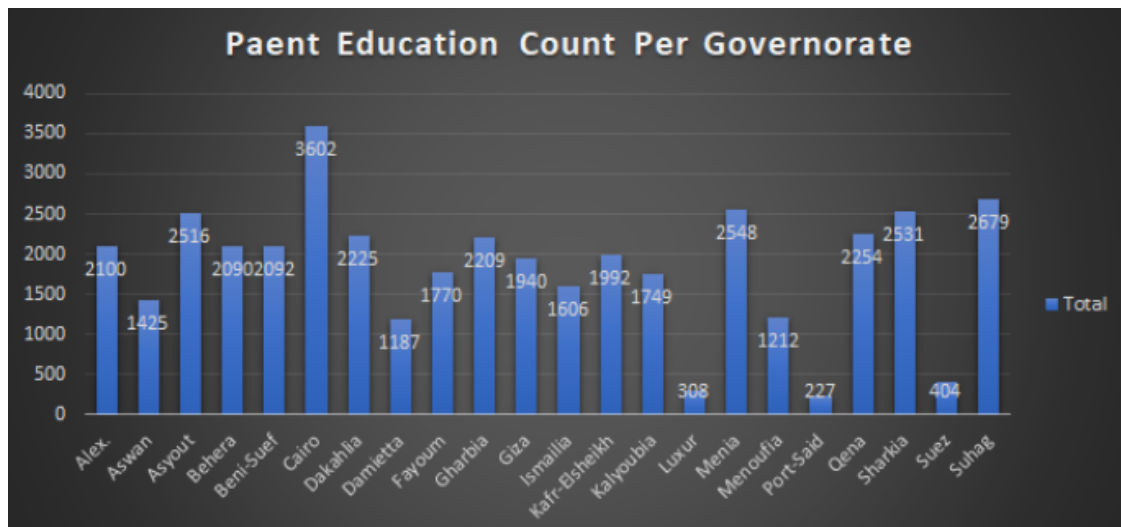
**A few FACTS arise here**:

1. More than 8,000 data point don't have entry for parent education "**missing**". Which was outlined earlier in the Data Wrangling section.

2. Cairo Was the **highest** governorate in educated parents count, with 3602 parent.

3. Port-Said again was the **lowest** governorate in educated parents count, with only 227 parent.

### 4.3.4   Education Total (Parent and Individuals) Vs Governorate

In this section we compare both attributes **educ** "execluding illetrate" and **fteducst** for each one of the governorates. We present both a table and chart for the following comparison.

In the following table we compare the total number of educated (fathers and Individuals) per governorate.

```
[ ]: Image("drive/My Drive/Economics Data Analysis Competition/Images/
     ↪PT_ALLEducGovTable.png")
```
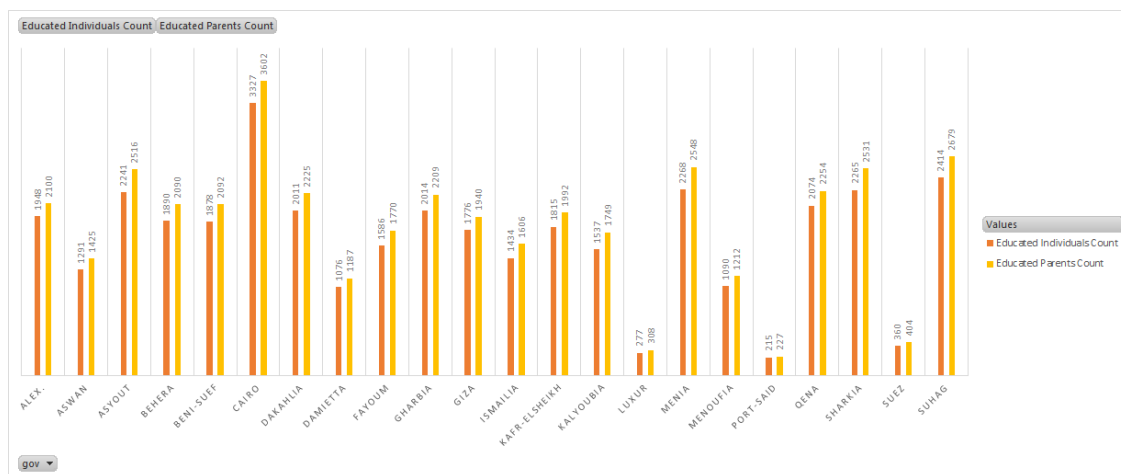
[ ]:

| 3 | Governorate ▼ | Parents Educated Count | Educated Individuals Count |
|---|---|---|---|
| 4 | Alex. | 2100 | 1948 |
| 5 | Aswan | 1425 | 1291 |
| 6 | Asyout | 2516 | 2241 |
| 7 | Behera | 2090 | 1890 |
| 8 | Beni-Suef | 2092 | 1878 |
| 9 | Cairo | 3602 | 3327 |
| 10 | Dakahlia | 2225 | 2011 |
| 11 | Damietta | 1187 | 1076 |
| 12 | Fayoum | 1770 | 1586 |
| 13 | Gharbia | 2209 | 2014 |
| 14 | Giza | 1940 | 1776 |
| 15 | Ismailia | 1606 | 1434 |
| 16 | Kafr-Elsheikh | 1992 | 1815 |
| 17 | Kalyoubia | 1749 | 1537 |
| 18 | Luxur | 308 | 277 |
| 19 | Menia | 2548 | 2268 |
| 20 | Menoufia | 1212 | 1090 |
| 21 | Port-Said | 227 | 215 |
| 22 | Qena | 2254 | 2074 |
| 23 | Sharkia | 2531 | 2265 |
| 24 | Suez | 404 | 360 |
| 25 | Suhag | 2679 | 2414 |
| 26 | **Grand Total** | **40666** | **36787** |

This chart plots the previous table, but in chart-form.

```
Image("drive/My Drive/Economics Data Analysis Competition/Images/
↪PT_ALLEducGovChart.png")
```



**A few FACTS arise here**:

1. Cairo Was the **highest** governorate in **both** educated fathers count and total educated individuals count, with 3602 educated fathers and 3327 educated individuals.

2. Port-Said again was the **lowest** governorate in educated fahters count, with only 227 educated fathers.

### 4.3.5 Marital (Parent and Individuals) Vs Governorate

In this section we compare the Martial attribute with its types or different values for each governorate. We present both a table and chart for the following comparison.
Marital Status can be one of the following 6 cases (according to the dataset):

1. Contractually Married

2. Divorced

3. Less Than Miniumum Age

4. Married

5. Never Married

6. Widowed(er)

In the following table we compare number of individuals of each marital status group per governorate.

```
[ ]:  Image("drive/My Drive/Economics Data Analysis Competition/Images/maritalGovPie.
      ↪png")
```

[ ]:

| | Count of marital | Marital Status | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | Governorate | contractually married | divorced | less than minimum age | married | never married | widowed(er) | Grand Total |
| 5 | Alex. | 5 | 31 | 784 | 1109 | 409 | 126 | 2464 |
| 6 | Aswan | | 13 | 612 | 674 | 316 | 80 | 1695 |
| 7 | Asyout | 10 | 20 | 1331 | 1222 | 401 | 119 | 3103 |
| 8 | Behera | 7 | 16 | 919 | 1185 | 307 | 118 | 2552 |
| 9 | Beni-Suef | | 17 | 1077 | 1062 | 287 | 118 | 2561 |
| 10 | Cairo | 5 | 64 | 1231 | 1765 | 734 | 287 | 4086 |
| 11 | Dakahlia | 2 | 20 | 1002 | 1210 | 287 | 133 | 2654 |
| 12 | Damietta | 2 | 14 | 569 | 653 | 171 | 60 | 1469 |
| 13 | Fayoum | 9 | 15 | 939 | 920 | 243 | 89 | 2215 |
| 14 | Gharbia | 2 | 16 | 989 | 1228 | 323 | 120 | 2678 |
| 15 | Giza | 5 | 28 | 928 | 1061 | 269 | 106 | 2397 |
| 16 | Ismailia | 4 | 16 | 766 | 894 | 214 | 73 | 1967 |
| 17 | Kafr-Elsheikh | 3 | 15 | 915 | 1049 | 291 | 110 | 2383 |
| 18 | Kalyoubia | 1 | 15 | 872 | 962 | 211 | 104 | 2165 |
| 19 | Luxur | | 2 | 140 | 166 | 50 | 18 | 376 |
| 20 | Menia | 1 | 18 | 1348 | 1278 | 358 | 118 | 3121 |
| 21 | Menoufia | 2 | 16 | 576 | 659 | 152 | 68 | 1473 |
| 22 | Port-Said | 4 | 4 | 68 | 124 | 46 | 16 | 262 |
| 23 | Qena | 2 | 27 | 1050 | 1087 | 419 | 130 | 2715 |
| 24 | Sharkia | 5 | 18 | 1208 | 1358 | 342 | 122 | 3053 |
| 25 | Suez | 1 | 7 | 157 | 221 | 67 | 31 | 484 |
| 26 | Suhag | 2 | 13 | 1413 | 1322 | 429 | 134 | 3313 |
| 27 | **Grand Total** | **72** | **405** | **18894** | **21209** | **6326** | **2280** | **49186** |

This chart plots the previous table, but in chart-form.

```
[ ]: Image("drive/My Drive/Economics Data Analysis Competition/Images/
     ↪maritalGovTable.png")
```

[ ]:
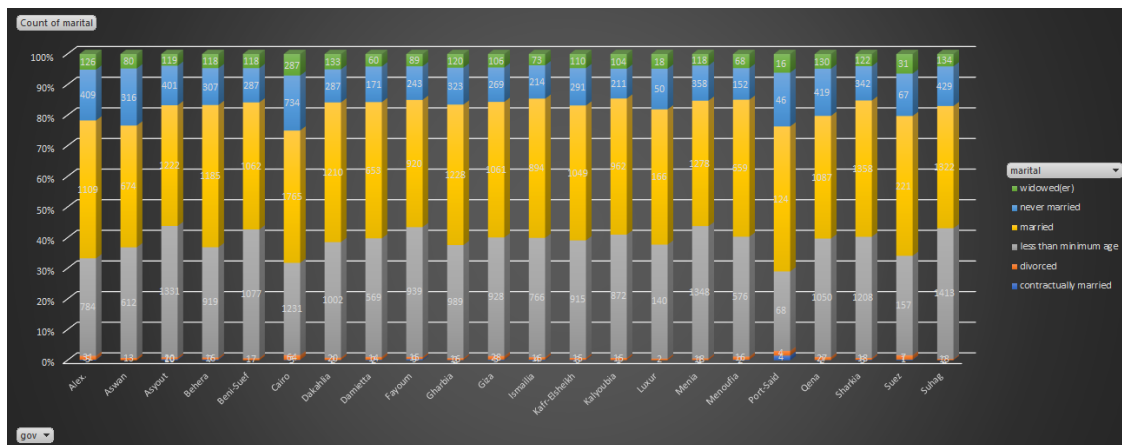


**A few FACTS arise here:** 1. **Marriages Count:** Cairo has the **highest** marriage count in all governorates with 1765 marriages.

2. **Divorces Count:** Divorces count is low. But count of divorce is very simliar in all governorates except for Cairo which is the **highest** with approximately 64 divorces which is more than **triple or double** all other governorates.

3. **Widow and Never Married Count:** Again Cairo is the highest governorate in numbers of widows and never married individauls, with counts equal to 287 and 734 respectively.

### 4.3.6  Sector Of Activity Vs Governorate

In this section we compare the Sector of Activity attribute i.e **crecac1d** attribute with its different types or different activities for each governorate. We present both a table and chart for the following comparison.

Sector Of Activity of an individual can be one of the following 21 cases (according to our dataset):

1. A:Agriculture, forestry and fishing

2. B:Mining and quarrying

3. C:Manufacturing

4. D:Electricity,gas,steam and air conditioning supply

5. E:Water supply;sewage,waste management and remediation activities

6. F:Construction

7. G:Wholesale and retail trade; repair of motor vehicles and motorcycles

8. H:Transportation and storage

19

9. I:Accomodation and food service activities

10. J:Information and communication

11. K:Financial and insurance activities

12. L:Real estate activities

13. M:Professional, scientific and technical activities

14. N:Administrative and support service activities

15. O:Public administration and defense; compulsory social security

16. P:Education

17. Q:Human health and social work activities

18. R:Arts, entertainment and recreation

19. S:other service activities

20. T:Activities of extraterritorial organizations and bodies

21. U:Activities of extraterritorial organizations and bodies

**Yup, this is too much to be put in one table or one graph**
So I will divide the 21 sectors of activities into 4 groups of 5 activities each. Such that I will compare each group to all governorates, and then get insights per group. For each group I draw a number of facts from each chart.
**The main insight I draw is the governorate(s) which has the highest number of individuals working in that sector of activity.**

**First GROUP (A - E)** In the following table we compare the **FIRST FIVE** sectors of activity per governorate.

```
[ ]: Image("drive/My Drive/Economics Data Analysis Competition/Images/
     ↪firstGroupTable.png")
```

[ ]:

| Count of crecac1d | Sector of Activity | | | | | |
|---|---|---|---|---|---|---|
| Governorate | A:Agriculture, forestry and fishing | B:Mining and quarrying | C:Manufacturing | D:Electricity | E:Water supply | Grand Total |
| Alex. | 21 | 1 | 141 | 16 | 5 | 184 |
| Aswan | 205 | 7 | 17 | 8 | 6 | 243 |
| Asyout | 309 | 1 | 64 | 1 | 3 | 378 |
| Behera | 345 | 1 | 53 | 7 | 6 | 412 |
| Beni-Suef | 396 | | 48 | 1 | 3 | 448 |
| Cairo | 15 | 3 | 212 | 16 | 14 | 260 |
| Dakahlia | 218 | 1 | 133 | 4 | 5 | 361 |
| Damietta | 51 | | 132 | 1 | 14 | 198 |
| Fayoum | 358 | 2 | 45 | 2 | 6 | 413 |
| Gharbia | 411 | | 130 | 1 | 5 | 547 |
| Giza | 231 | 2 | 141 | 4 | 5 | 383 |
| Ismailia | 182 | 1 | 72 | 18 | 9 | 282 |
| Kafr-Elsheikh | 448 | | 35 | 4 | 5 | 492 |
| Kalyoubia | 160 | | 129 | 2 | 11 | 302 |
| Luxur | 26 | | 3 | | 1 | 30 |
| Menia | 409 | 2 | 61 | 3 | 6 | 481 |
| Menoufia | 160 | 2 | 75 | 7 | 2 | 246 |
| Port-Said | 2 | | 19 | | | 21 |
| Qena | 249 | 2 | 33 | 3 | 7 | 294 |
| Sharkia | 401 | 1 | 123 | 7 | 3 | 535 |
| Suez | 7 | 1 | 34 | 5 | 1 | 48 |
| Suhag | 360 | 5 | 37 | 2 | 3 | 407 |
| Grand Total | 4964 | 32 | 1737 | 112 | 120 | 6965 |

This chart plots the previous table, but in chart-form.

```
[ ]: Image("drive/My Drive/Economics Data Analysis Competition/Images/
    ↪firstGroupChart.png")
```

[ ]:
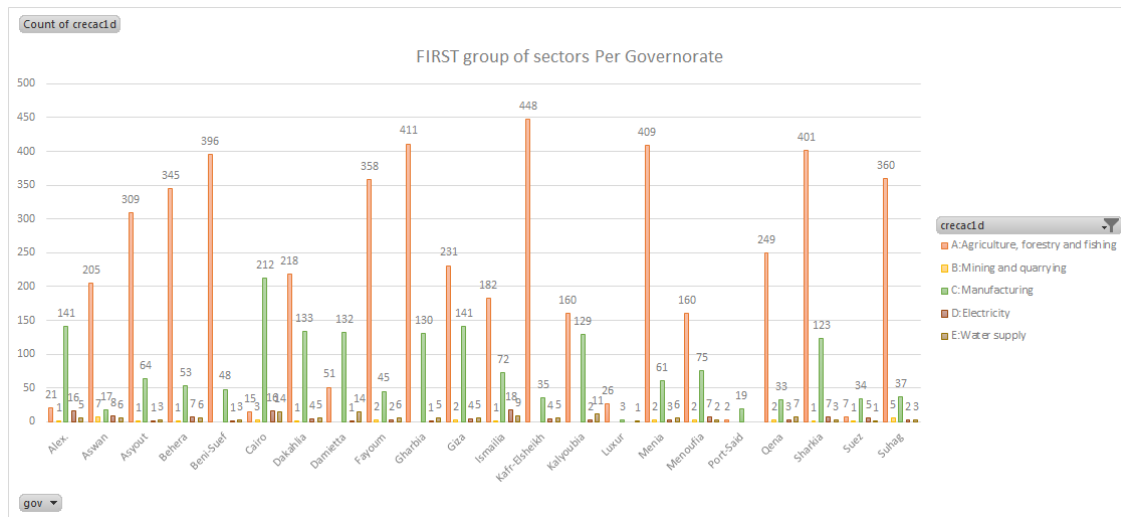


**A few FACTS arise here:**

**By Highest I mean the highest governorate in the number of individuals working in that respective sector.**

1. Agriculture, forestry and fishing: Kafr-Elsheikh was the **highest** followed by Gharbia, Menia and Qena.

2. B:Mining and quarrying: Aswan was the **highest** followed by Sohag.

3. C:Manufacturing: Cairo was the **highest** governorate.

4. D:Electricity,gas,steam and air conditioning supply: Ismailia was the **highest** followed by Cairo and Alexandria.

5. E:Water supply,sewage,waste management and remediation activities: The **highest two** governorates were Damietta and Cairo.

**Second GROUP (F - J)** In the following table we compare the SECOND FIVE sectors of activity per governorate.

```
[ ]: Image("drive/My Drive/Economics Data Analysis Competition/Images/
    ↪secondGroupTable.png")
```

[ ]:

| Count of crecac1d | Sector of Activity | | | | | |
|---|---|---|---|---|---|---|
| Governorate | F:Construction | G:Wholesale and retail | H:Transportation | I:Accomodation | J:Information | Grand Total |
| Alex. | 73 | 151 | 75 | 30 | 9 | 338 |
| Aswan | 50 | 61 | 25 | 15 | 5 | 156 |
| Asyout | 111 | 120 | 42 | 11 | 7 | 291 |
| Behera | 60 | 129 | 51 | 17 | 10 | 267 |
| Beni-Suef | 113 | 104 | 41 | 22 | 4 | 284 |
| Cairo | 109 | 230 | 91 | 37 | 44 | 511 |
| Dakahlia | 69 | 129 | 83 | 14 | 5 | 300 |
| Damietta | 38 | 49 | 31 | 9 | 2 | 129 |
| Fayoum | 92 | 91 | 38 | 17 | 1 | 239 |
| Gharbia | 67 | 130 | 58 | 27 | 1 | 283 |
| Giza | 64 | 110 | 59 | 15 | 12 | 260 |
| Ismailia | 57 | 103 | 65 | 18 | 9 | 252 |
| Kafr-Elsheikh | 62 | 113 | 43 | 11 | 4 | 233 |
| Kalyoubia | 56 | 117 | 55 | 21 | 10 | 259 |
| Luxur | 18 | 7 | 4 | 3 | 1 | 33 |
| Menia | 91 | 104 | 51 | 24 | 8 | 278 |
| Menoufia | 43 | 56 | 21 | 8 | 4 | 132 |
| Port-Said | 4 | 10 | 15 | 3 | 2 | 34 |
| Qena | 157 | 55 | 35 | 26 | 4 | 277 |
| Sharkia | 99 | 121 | 55 | 20 | 4 | 299 |
| Suez | 8 | 26 | 20 | 4 | 3 | 61 |
| Suhag | 137 | 96 | 52 | 23 | 6 | 314 |
| Grand Total | 1578 | 2112 | 1010 | 375 | 155 | 5230 |

This chart plots the previous table, but in chart-form.

```
[ ]: Image("drive/My Drive/Economics Data Analysis Competition/Images/
     ↪secondGroupChart.png")
```

[ ]:
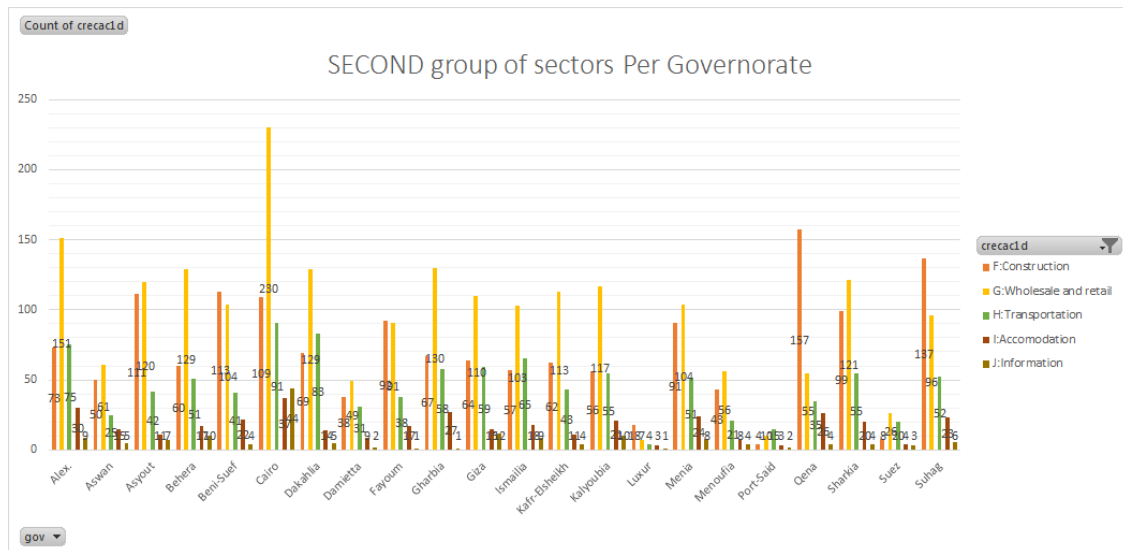


**A few FACTS arise here:**

6. F:Construction: The **highest** governorate was Qena then followed by Sohag

7. G:Wholesale and retail trade, repair of motor vehicles and motorcycles: The **highest** governorate was Cairo with almost the double of all other governorates.

8. H:Transportation and storage: The **highest** governorate was Cairo followed by Dakahlia.

9. I:Accomodation and food service activities: The **highest** governorate was Cairo followed by Alexandria.

10. J:Information and communication: The **highest** governorate was Cairo with a significant difference than other governorates that almost reaches **4x** the other governorates.

**Third GROUP (K - O)** In the following table we compare the THIRD FIVE sectors of activity per governorate.

```
[ ]: Image("drive/My Drive/Economics Data Analysis Competition/Images/
     ↪thirdGroupTable.png")
```

[ ]:

| Count of crecac1d | Sector of Activity | | | | | |
|---|---|---|---|---|---|---|
| Governorate | K:Financial and insurance | L:Real estate | M:Professional, scientific | N:Administrative | O:Public administration | Grand Total |
| Alex. | 7 | 1 | 14 | 2 | 45 | 69 |
| Aswan | 4 | | 9 | 2 | 57 | 72 |
| Asyout | 4 | 2 | 9 | 3 | 52 | 70 |
| Behera | 5 | | 6 | 4 | 80 | 95 |
| Beni-Suef | 3 | | 10 | 1 | 76 | 90 |
| Cairo | 28 | 3 | 59 | 28 | 117 | 235 |
| Dakahlia | 4 | | 7 | 6 | 38 | 55 |
| Damietta | 3 | | 11 | 3 | 27 | 44 |
| Fayoum | 6 | | 4 | 1 | 59 | 70 |
| Gharbia | 2 | | 12 | 2 | 66 | 82 |
| Giza | 11 | 2 | 17 | 17 | 41 | 88 |
| Ismailia | 13 | | 11 | 1 | 71 | 96 |
| Kafr-Elsheikh | 4 | | 8 | 4 | 60 | 76 |
| Kalyoubia | 8 | | 5 | 10 | 39 | 62 |
| Luxur | 1 | | 3 | 2 | 13 | 19 |
| Menia | 6 | | 4 | 1 | 74 | 85 |
| Menoufia | 1 | | 12 | 4 | 41 | 58 |
| Port-Said | 2 | | 2 | 2 | 24 | 30 |
| Qena | 3 | | 4 | 3 | 61 | 71 |
| Sharkia | 3 | | 8 | 5 | 82 | 98 |
| Suez | 3 | | 3 | 1 | 10 | 17 |
| Suhag | 10 | 1 | 6 | 2 | 40 | 59 |
| **Grand Total** | **131** | **9** | **224** | **104** | **1173** | **1641** |

This chart plots the previous table, but in chart-form.

```
[ ]: Image("drive/My Drive/Economics Data Analysis Competition/Images/
     ↪thirdGroupChart.png")
```

[ ]:
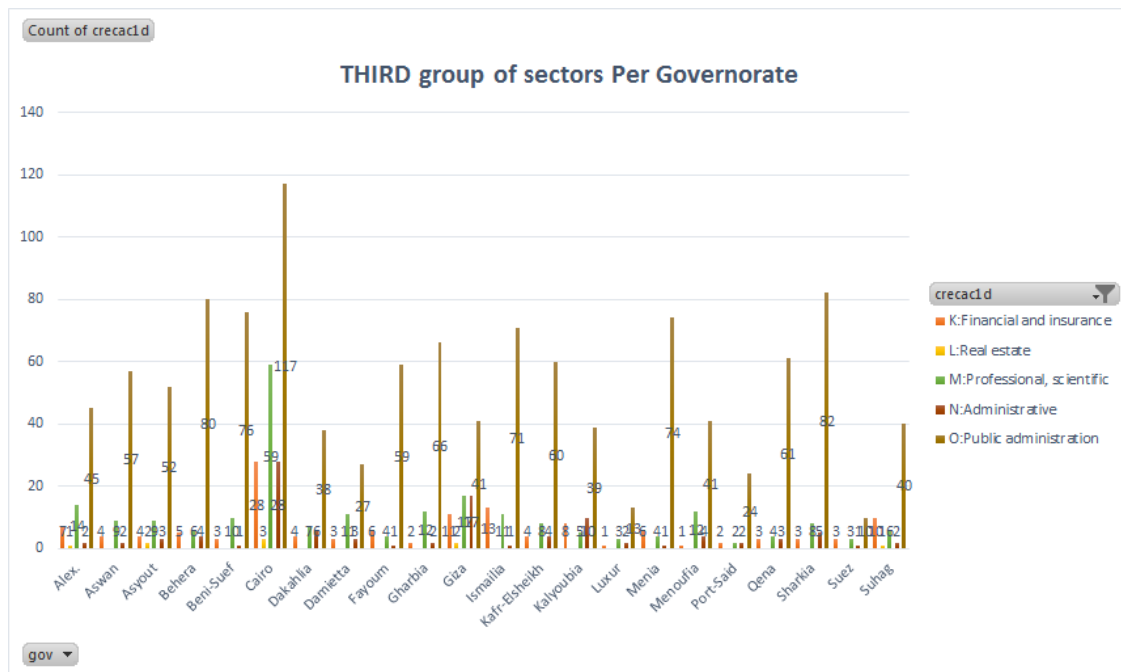
**A few FACTS arise here:**

11. K:Financial and insurance activities: The **highest** governorate was Cairo.

12. L:Real estate activities: The data points is very low in this category byt the **highest** governorate was Cairo .

13. M:Professional, scientific and technical activities: The **highest** governorate was Cairo.

14. N:Administrative and support service activities: The **highest** governorate was Cairo followed by Dakahlia.

15. O:Public administration and defense; compulsory social security: The **highest** governorate was Cairo.

**Fourth GROUP (P - U)**    In the following table we compare the FOURTH FIVE sectors of activity per governorate.

```
[ ]: Image("drive/My Drive/Economics Data Analysis Competition/Images/
     ↪fourthGroupTable.png")
```

[ ]:

| Count of crecac1d | Sector of Activity | | | | | | |
|---|---|---|---|---|---|---|---|
| Governorate | P:Education | Q:Human health | R:Arts, entertainment | S:other service activities | T:Extraterritorial organizations | U:Extraterritorial organizations | Grand Total |
| Alex. | 64 | 31 | 5 | 28 | 4 | | 132 |
| Aswan | 56 | 14 | 2 | 7 | | | 79 |
| Asyout | 80 | 30 | 2 | 18 | 2 | | 132 |
| Behera | 80 | 29 | 2 | 22 | | | 133 |
| Beni-Suef | 56 | 16 | 6 | 15 | 6 | | 99 |
| Cairo | 104 | 53 | 11 | 39 | 19 | 2 | 228 |
| Dakahlia | 74 | 41 | 4 | 22 | 1 | | 142 |
| Damietta | 44 | 27 | 1 | 8 | | | 80 |
| Fayoum | 79 | 5 | 1 | 9 | 1 | | 95 |
| Gharbia | 78 | 27 | 1 | 24 | 1 | | 131 |
| Giza | 56 | 20 | 10 | 15 | 4 | 2 | 107 |
| Ismailia | 85 | 23 | | 19 | | | 127 |
| Kafr-Elsheikh | 96 | 13 | 4 | 22 | 2 | | 137 |
| Kalyoubia | 72 | 15 | 2 | 15 | 4 | | 108 |
| Luxur | 17 | 3 | 1 | 1 | | | 22 |
| Menia | 125 | 21 | 5 | 17 | | | 168 |
| Menoufia | 69 | 20 | 5 | 10 | 2 | | 106 |
| Port-Said | 18 | 3 | | 2 | | | 23 |
| Qena | 74 | 17 | 2 | 5 | 1 | | 99 |
| Sharkia | 113 | 31 | 9 | 19 | 5 | | 177 |
| Suez | 20 | 4 | | 2 | | | 26 |
| Suhag | 88 | 14 | 2 | 15 | 1 | | 120 |
| Grand Total | 1548 | 457 | 75 | 334 | 53 | 4 | 2471 |

This chart plots the previous table, but in chart-form.

```
[ ]: Image("drive/My Drive/Economics Data Analysis Competition/Images/
     ↪fourthGroupChart.png")
```

[ ]:
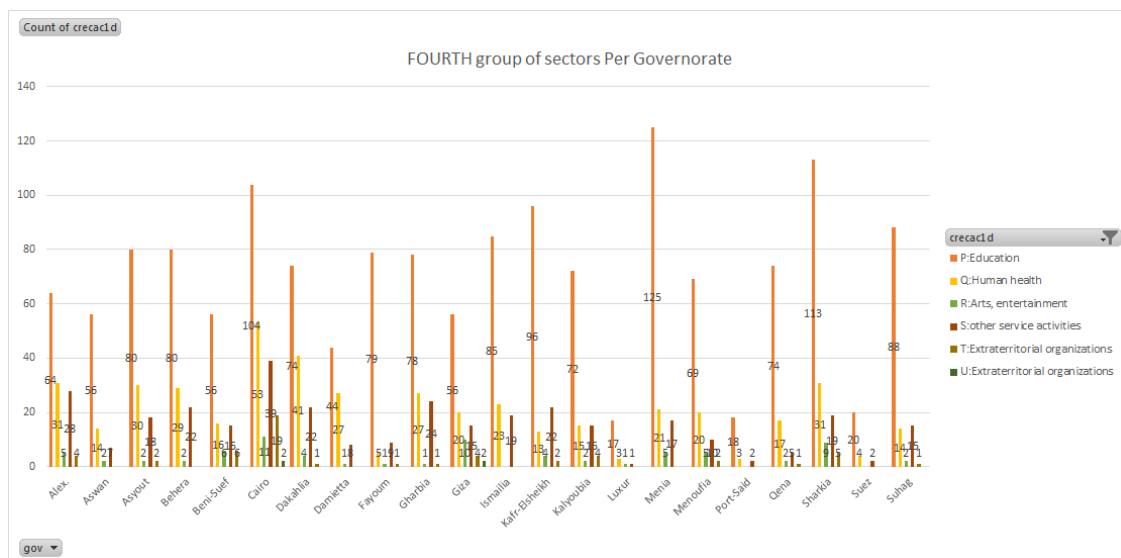


**A few FACTS arise here:**

16. P:Education: The **highest** governorate was Menia followed by Sharkia.

17. Q:Human health and social work activities: The **highest** governorate was Cairo followed by Dakahlia.

18. R:Arts, entertainment and recreation: The **highest** governorate was Cairo followed by Gize with one a **one-individual** difference.

25

19. S:other service activities: The **highest** governorate was Cairo .

20. T:Activities of extraterritorial organizations and bodies: The **highest** governorate was Cairo.

21. U:Activities of extraterritorial organizations and bodies: The number of individuals collected for this sector was **very low**. Only two governorates had individuals working in this sector. Cairo and Giza with two individuals count.

## 4.4 Research Question 4 (Present Facts about each governorate in terms of Marital Status as a share of the population. (Population data can be exported from the CAPMAS website, accessible at the library).)

This research question was addressed in the previous

## 4.5 Research Question 5 (Any relationship between Father Education and Individual Education, per governorate in Egypt? Search about the economic theory that can support your findings and present it.)

### 4.5.1 Investigating Insights

In the following table we compare the total number of educated (fathers and Individuals) per governorate.

```
[ ]: Image("drive/My Drive/Economics Data Analysis Competition/Images/
      ↪PT_ALLEducGovTable.png")
```
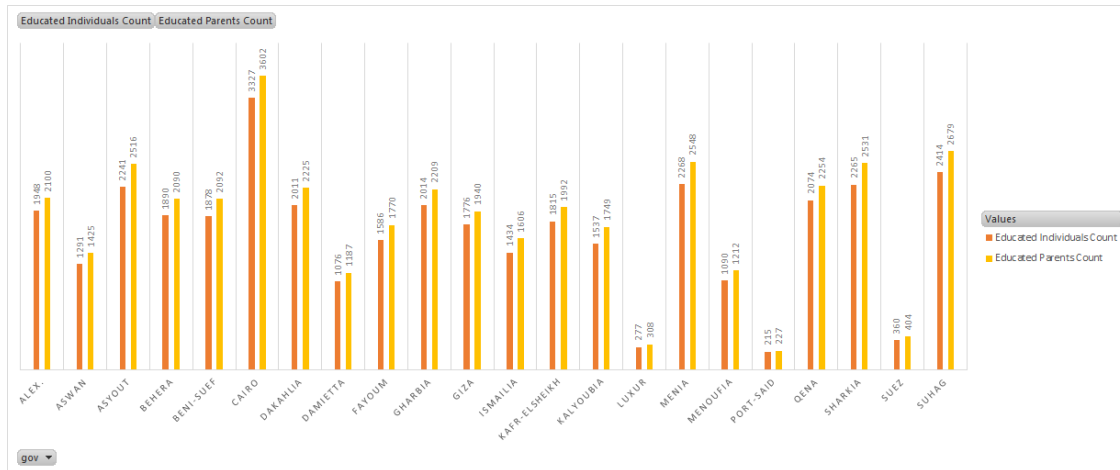
[ ]:

| | Governorate | Parents Educated Count | Educated Individuals Count |
|---|---|---|---|
| 3 | | | |
| 4 | Alex. | 2100 | 1948 |
| 5 | Aswan | 1425 | 1291 |
| 6 | Asyout | 2516 | 2241 |
| 7 | Behera | 2090 | 1890 |
| 8 | Beni-Suef | 2092 | 1878 |
| 9 | Cairo | 3602 | 3327 |
| 10 | Dakahlia | 2225 | 2011 |
| 11 | Damietta | 1187 | 1076 |
| 12 | Fayoum | 1770 | 1586 |
| 13 | Gharbia | 2209 | 2014 |
| 14 | Giza | 1940 | 1776 |
| 15 | Ismailia | 1606 | 1434 |
| 16 | Kafr-Elsheikh | 1992 | 1815 |
| 17 | Kalyoubia | 1749 | 1537 |
| 18 | Luxur | 308 | 277 |
| 19 | Menia | 2548 | 2268 |
| 20 | Menoufia | 1212 | 1090 |
| 21 | Port-Said | 227 | 215 |
| 22 | Qena | 2254 | 2074 |
| 23 | Sharkia | 2531 | 2265 |
| 24 | Suez | 404 | 360 |
| 25 | Suhag | 2679 | 2414 |
| 26 | Grand Total | 40666 | 36787 |

This chart plots the previous table, but in chart-form.

```
Image("drive/My Drive/Economics Data Analysis Competition/Images/
↪PT_ALLEducGovChart.png")
```

[ ]:



**We have presented earlier facts about this plot** But now we will see the relation between parents education count and individuals education count.

From the graph I can see clearly that governorates with **High Fathers Education** are **High in Individuals Education** count.

Therefore, I can see that Fathers education is **directly proportional** to Individuals education.

### 4.5.2 Supporting Economical Theory and references

As shown in references [4], [5], and [6] all researches have successfully concluded that parent's educational level positively affects their childrens' educational level. As stated by [4] "College-educated parents spend more time with their children, model achievement-oriented behavior, provide opportunities for their children to engage in achievement-oriented experiences, engage in ageappropriate activities, and cultivate their children's talents. These behaviors, generally, lead to a child's educational success.".

In [5] they conclude with the following results "we found that parental education affects children's aspirations for their own education as well as their actual educational achievement through adolescence." . Other research (e.g., Alexander, Entwisle, & Bedinger, 1994; Davis-Kean, 2005; Klebanov, Brooks-Gunn, & Duncan, 1994; Smith et al., 1997) has shown that parental education is linked to the parents providing a more stimulating physical, cognitive, and emotional environment in the home, and more accurate beliefs about their children's actual achievement. These proximal processes likely affect the developing child's achievement-related aspirations and actual achievement behavior.

27

In the study referenced under [6] they have the following conclusions: "The findings and results obtained from the analysis of the study revealed the importance of educated parents and its relationship with school related decisions of parents. It was concluded that parental education was considered necessary for the proper grooming of children. The decisions of highly educated parents are more relevant and reflective as compared to the low and average level educated parents. Further, highly educated parents have positive and significant relation with students' subject selection. Furthermore, based on the results it was concluded that there is a large correlation between parents' education and students' academic achievement."

# 5 Limitations and References

## 5.1 Limitations

- The tests done in this notebook are yet tentative ones, as we didn't use any Machine learning or inferential statisctics which would of made things much accurate and better relations.

- Also there are a number of attributes which include a largen number of missing data points, as mentioned under the Data Wrangling section. All of the conclusions pointed out through this analysis were including this missing data. We **didn't try to imputate** the missing data with average values, zeros or anything.

## 5.2 References

1. http://www.erfdataportal.com/index.php/catalog/157

2. https://catalog.ihsn.org/index.php/catalog/5869

3. https://izajold.springeropen.com/articles/10.1186/2193-9020-2-8#Sec5

4. https://militaryfamilies.psu.edu/wp-content/uploads/2020/01/Parents-Educational-Levels-Influence-on-Child-Educational-Outcomes.20Jan06.final_.pdf

5. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2853053/

6. https://www.researchgate.net/publication/348138126_Relationship_between_Parents'_Education_and_th

**This project was done by Youssef M. Hussien as a submission to the Economic Association Competition Data Analysis.**