****

**PROFILE - FREELANCER**

EduManta ( <https://www.edumanta.com/>) is an Online Educational Platform that provides live courses to professionals and the other wing of EduManta helps college students across the world by providing them doubts clarification and assignment solutions.

So, we are connecting with you to help those students in their academic doubts and assignments. Also, please note you will be hired as a Freelancer.

**Job Profile**

From the August; the session will start in our targeted countries and you will be required to provide the solution for assignments as and when needed. As assignments from these developed countries are real-time, and based on real situations, it will provide you with ample opportunity to learn. Not only that, you will get real-time feedback from the USA’s university professors and all this experience will not only help you to earn a good amount per assignment or doubt clearing session but also an enriching experience that will help you to crack your future jobs.

Further on connecting with us will provide you a lifetime opportunity to earn part-time income from what you have learned throughout your life. The payment will be made on per assignment basis.

If you are interested in connecting with us, do solve this assignment and send it via intershala chat or via email.

**SELECTION PROCESS – PYTHON ASSIGNMENT**

# Assessment Requirements / Tasks (include all guidance notes)

This assignment will use [employment data](https://statswales.gov.wales/Catalogue/Business-Economy-and-Labour-Market/People-and-Work/Employment/Jobs/Whole-Workforce/workplaceemployment-by-industry-area) of Wales from the StatsWales data source. This dataset provides workplace employment estimates, or estimates of total jobs, for Wales and its NUTS2 areas, along with comparable UK data disaggregated by industry section.

For this assignment students will undertake a data analysis and machine learning approach to reveal the workplace employment landscape of Wales.

# Data processing

# Download the dataset for the period 2009 – 2018 and create a data frame that concatenates Wales (total) employment value only.

# Check for any null value or outlier. If found replace that with mean value.

# Change the name of the industries as bellow

The final data frame should look like following

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Industry | 2009 | 2010 | 2011 | 2012 | 2013 | 2014 | 2015 | 2016 | 2017 | 2018 |
| Agriculture |  |  |  |  |  |  |  |  |  |  |
| Production |  |  |  |  |  |  |  |  |  |  |
| Construction |  |  |  |  |  |  |  |  |  |  |
| Retail |  |  |  |  |  |  |  |  |  |  |
| ICT |  |  |  |  |  |  |  |  |  |  |
| Finance |  |  |  |  |  |  |  |  |  |  |
| Real Estate |  |  |  |  |  |  |  |  |  |  |
| Professional Service |  |  |  |  |  |  |  |  |  |  |
| Public\_Adminstration |  |  |  |  |  |  |  |  |  |  |
| Other Service |  |  |  |  |  |  |  |  |  |  |

# Data analysis

For each question provide graph/chart along with your own interpretation (~ 50 words)

* 1. Which industry employed highest and lowest workers over the period?
  2. Which industry has the highest and lowest overall growth over the period?
  3. Which years are the best and worst performing year in relation to number of employments? (highest and lowest employment)

# Visual analysis

Create a dynamic scatter/bubble plot showing the change of workforce number over the period using [Plotly express](https://plot.ly/python/plotly-express/).

# PCA/Correlation

* 1. Undertake a PCA (PC=2; columns should be like PC1, PC2, Industry) and produce a scatter plot. Write your interpretation about the plot and in relation to the analysis of section 2 & 3 (for example which industries are correlated over the years as well as in PCA etc.)
  2. Make a year wise correlation for each industry. Does the aforementioned industries are also correlated over the years? Explain your answer.

1. **Clustering (k means & hierarchical)**
   1. Using the best and worst performing year column’s employment data (2.3) undertake a K means clustering analysis (K=2 & 3) and identify industries cluster together. Write your own interpretation (~100 words).
   2. Using the same dataset (best & worst performing) create a hierarchical cluster. Compare the cluster with k means clusters.
2. **Discussion**

Provide a brief discussion (~ 300 words) on employment landscape of Wales based on the employment data analysis results.

# Assessment Criteria

|  |  |
| --- | --- |
| 1.1 Data preparation | **05** |
| 1.2 Data preparation | **05** |
| 1.3 Data preparation | **05** |
| 2.1 Data analysis | **05** |
| 2.2 Data analysis | **05** |
| 2.3 Data analysis | **05** |
| 3 Visual analysis | **20** |
| 4.1 PCA | **10** |
| 4.1 Correlation | **10** |
| 5.1 Clustering | **10** |
| 5.2 Clustering | **10** |
| 6 Discussion | **10** |

**Note:**

* **Download the Document and Rename it as ( YourName\_Python\_EduManta)**
* **Deadline- 29th May, 2020**
* **You can provide part solution or complete solution as per your knowledge.**