London Workforce Mosaic

Understanding Industry-specific trends in the Employment Market

Rahul Mehta

MSc Data Analytics

Student Id: 22048235

May 2024

This dissertation is submitted in part for the fulfilment of the degree of MSc Data Analytics for London Metropolitan University

ABSTRACT

The primary objective of creating this dissertation is to create models that is going to predict the number of jobs in London and UK. The research has been studied by employing statistical techniques and machine learning algorithms. The data has been collected from *Office for National Statistics* website. The models being implemented for the forecasting are *Regression*, *Time Series and ARIMA*. The study is going to analyse and anticipate the upcoming workforce trends by getting the actual numbers from the models and compared them to analyse which one is more accurate and use the right information for strategic planning. The models are created and studied technically along with the step-by-step process of how they are implemented and how the information is drawn from them.

ACKNOWLEDGEMENT

I would like to express my gratitude to Professor Abdelkrim Ouazzane and Project supervisor Professor Sahar Al-Sudani for giving me an opportunity to work on dissertation. The valuable information, constructive feedback and fruitful suggestions eventually boosted my confidence to do the research analysis and brainstorming in a fruitful manner. Finally, I would like to thank to London Metropolitan University, for having me a part of this course which provided me an enormous knowledge that taught me doing on rigorous research, finding the key information, using of models for analysing. This dissertation helped me gaining some technical skills about the application of Power BI and Python Scripts for the development of machine learning models that eventually is going to shape my future in every aspect.

TABLE OF CONTENTS

CHAPTER – 1	**INTRODUCTION**	**PAGE**
1.1	Background & Context……	3
1.2	Research Question with Problem Statement……....	3
1.3	Goals of Investigation…….	3
1.4	Importance of Research…….	3-4
1.5	Scope & Constraints……...	4
CHAPTER – 2	**LITERATURE REVIEW**	
2.1	Introduction of Employment forecasting……	5
2.2	Theoretical structure……	5
2.3	Earlier Research Study……	5-6
2.4	Conceptual framework……	6
2.5	Gaps in existing literature……...	6
2.6	How my data analysis fills these gaps…….	6-7
CHAPTER – 3	**RESEARCH METHODOLOGY (DATA ANALYSIS LIFE CYCLE**	
3.1	Discovery…….....	7-8
3.2	Data Preparation…….	8-10
3.3	Plan Model…….	10-11
3.4	Model Building……...	11-18
3.5	Communication result……	18-20
3.6	Operationalize…….	20

CHAPTER-4	**BUILDING MACHINE LEARNING MODEL USING PYTHON SCRIPT**	**PAGE**
4.1	About Python……	20
4.2	Python merits……	20
4.3	Connecting Python with Power BI desktop……..	21-24
4.4	Creating visuals in Power BI…….	24-26
4.5	Building machine learning model – ARIMA……..	27-31
CHAPTER-5	**TREND ANALYSIS**	
5.1	Trend 1 - 12 ……	32-28
CHAPTER-6	**COMPARTIVE STUDY**	
6.1	Study 1 – 6 ……	39-42
6.2	Limitations……	43
6.3	Future Research directions……..	43
CHAPTER-7	**CONCLUSION**	
7.1	Summary of Insights……...	43
7.2	Practical implications with contribution to knowledge…….	43
7.3	In Nutshell……	44
REFERENCES	……..	44-46

CHAPTER 1: INTRODUCTION

1.1 Background and Context

Employment or job generation is an important aspect of economy. This paper focuses on forecasting the number of jobs generated by different sectors of industry in London and UK. Having an ample number of jobs in all sectors of the industry results in a positive impact on the economy. There is a Keynesian multiplier theory which states that there is boost in the financial activity then it will eventually generate more revenue for the companies, creates employment opportunities with the availability of huge number of resources. Moreover, if the government injects some finance, then there would be a hike in the GDP(GANTI, 2019).

The information of the forecasted employment data provides valuable insights. These insights are going to facilitate the policy makers, business owners and government to create education, awareness and inductions as per the predicted data. It provides information about which sector is going to have emerging trend as far the job creation is concerned.

1.2 Research Question with problem statement

What reliable models for forecasting can be built for estimating potential job statistics in London and the UK spanning different sectors of the economy with accounting for variables like shifting demographics, economic trends, modifications to policies, and advancements in technology?

The research question caters the need for finding the valuable information by evaluating dataset for London and UK using modeling techniques such as Time series, Regression in Power BI and ARIMA by using Python script. By considering seriously several factors which impact labour dynamics, such as government policies, demographic shifts, advancements in technology, and macroeconomic instances, it bears focus to how tough the task is.

1.3 Goals of investigation

There are plenty of factors that predicts the number of jobs in London and UK which are:

- Building the data analysis models such as Time series and Regression. The models predict the number of jobs created in specific sectors and analyse their difference.
- Analysis the performance of the model which is being accurate based on historical dataset.
- Finding the key elements that determines the sudden high or downtrend as far as the job market is concerned.
- Generating valuable insights which could be used by the concerned agencies for drafting policies making and planning.
- Finding the sudden high and low trends along with the information for its occurrence.

1.4 Importance of research

Forecasting job in London and UK trends plays vital role for taking key decisions by strategy makers, entrepreneurs, people looking for job and investors in the market. There are plenty of reasons which are as such:

- **Assessment of employment situation:** It enables the policy makers and entrepreneurs to analyse the type of skills required in the future to fulfill the job market. Apart from that it also helps to assess whether the specific industry is going to boom or decline. UK has different work and skills conditions, advantages, issues and points of entry. Further, an array of additional offerings, such as health care, might affect employment and capacity results. Furthermore, employment and skill services can influence those additional services, such as studying that can improve one's physical and mental health. Large-scale changes like the pandemic, longer working lives, and improvements in technology are also having varying impacts on different demographic groups and geographically spots.

- **Portfolio strategy**: It facilitates the investor to take accurate investment decisions by looking at the workforce job trends. For instance, boom in construction sector jobs means the investor can put the money in construction stocks.
- **Establishment of policy:** The forecasting makes sure that the governments have a look at the sectors where there will be decline in employment. This makes the picture clear to fix the issues by helping those sectors.
- **Financial planning:** The forecasting of jobs guides the government officials to draft policies that is going help in planning of infrastructure development and revenue generation (Local Government Association, 2022).

1.5 Scope and Constraints of my study

The scope of my study mainly comprises of:

- *Goal:* The goal of my investigation is to predict the employment numbers across various sectors in *London* and *UK* overall and get valuable insights and data.
- *Software:* The software I used for the doing the analysis are Power BI, Python, Microsoft excel, Microsoft Word.
- *Contribution:* With the valuable insights, I would like to provide the required information to the administrators, government officials and policy makers to formulate new policies and do certain amendments to adhere the concern by looking at the forecasting data.

The constraints are the limitations which could make the research challenging. The project's size dictates the amount of money and time that are needed to achieve the project considering bigger tasks take additional funds to be completed. The vigilant of scope creep at all levels of the work and make every attempt to avoid it. By producing thorough schedules and obtaining authorization from everybody involved in the project earlier than starting of production might prevent boundary creep.

- ***Time***: Each stage of the project will have distinct time limitations thus time planning is vital to its performance. There can be ramifications if there is trying to delay the project schedule, such as extended due dates adjustments to the collaboration calendar, or decreased space for preparing for it. Some of the basic aspects of the time constraint are entire timeframe for the project, time put into the project, individual goals and objectives, Hours set up for design and the creation, the number that comprises project steps.
- ***Cost:*** Any financial constraints for the undertaking along with the project expenditure in all its components are samples of cost boundaries. Items that may influence revenue comprise proposal costs, wage for teammates, price of the amenities, cost of repair, equipment cost etc.
- ***Risk***: Business are anything unanticipated that could result in a work. There are certain positive project risks, but many of them are bad. For instance, an emerging technology could potentially grow into public while the undertaking is running. The technology at hand may let your project go swifter, or it could increase market rivalry and lower the intrinsic value of the completed item. Applying assessment of risks and risk administration strategies, one may discover hazards associated with the project while taking preventive action which can be Clarity is lacking, Expanded reach, Time limitations, technical issues, Insufficient outcomes, technical issues and Exorbitant pricing.
- ***Quality*:** The amount to which the results of your project live up to initial hopes is known as the project's quality. Since the completed project is the finished product of your action, every project limitation has an influence on it. Whilst there are certain components of a project that could give rise to poor integrity but are not usually related to expense, schedule, assets, risk, or scope, the project's execution is also its own restriction. Some of them are too many updates in project, lack of coding and analysis skills and less coordination.
- ***Resources:*** Although these project requirements are monetary in the environment, assets and cost boundaries are tightly associated with each other. Inappropriate allocation of assets can lead to project delays, higher costs, and lower overall project performance. Resources may include individuals, Instruments, Services and software (Team Asana, 2024)

CHAPTER 2: LITERATURE REVIEW

2.1 Introduction of employment forecasting

The scientific and artistic process of forecast is a crucial component that distinguishes out in the broad field of data science. This will be a fascinating exploration into the global field of making predictions about the future using current data trends, outcomes, and actions. Predicting plays a vital part in following through insightful information which shapes strategies and promotes overall efficiency as businesses and organizations count more and more on data to make decisions. Basically, forecasting is an approach of estimating potential events based on the examination of past data and patterns. Locating patterns and groups that may be utilized to generate reliable projections is the aim of projecting in data science, regardless of the subject matter stock the costs, the environment sales numbers, or actual medical outcomes. To discover underlying trends and cycles in data, several statistical examples, machine learning techniques, and statistical methods are used during the analysis. Data scientists can create models that accurately project results and incorporate the inherent intricacies of the data using the power of analytics (TechNikhil, 2024).

The London alone will list up 30 thousand more job openings in the coming years ahead. After the span of 10 years, there are many retirements and to fill the gaps, there would be 5.7 million jobs produced in London. On the other hand, UK will generate 34 million jobs. 62 % of the jobs will be created in skilled sector for the people who hold higher qualifications. London being the central hub has the advantage of attracting indigenous and global talent as it provides opportunity to all over the world. Between 2016 and 2026, there will be 2,210,000 people will be going to retire and 310,000 new jobs will be generated. (Higham, n.d.).

2.2 Theoretical structure

The theoretical structure of the research mainly describes about the objectives of this research which is highlighting the trends. Moreover, its primary objective is to find the predicting numbers of jobs in London and UK. This will give the insight to the concerned authorities to make some rules, inductions to meet the future requirements. It also aims to develop certain machine learning models which is going to provide evidence for the prediction data.

2.3 Earlier research study

In the research, the think tanks believed that in the future there would be fewer work hours and more leisure time. The internet and technology have changed the whole working scenario, the people can work anytime and from at any location. Tomorrow gets influenced by what is going on today. By predicting the jobs, it not only helps to get the desired number but also help to identifies the areas in which the education and training sector need to reform their content to fulfill future work needs (Will Higham, n.d.).

UK employment is expected to expand by 1.1% every year even though this will be surpassed in London (1.5%), the rest of the country (1.3%), the West Midlands (1.2%), and the eastern part of England (1.2%), according to estimates made by accountants at EY. Employment development is predicted to be smaller in all other areas than the UK altogether. In line with the Ernst & Young regional economic forecasts for the UK presented in February 2024, there would be a 1.9% improvement in the total cost generated (GVA) or the value generated through generating goods and services in the UK between 2024 and 2027. This increase is anticipated to contribute to expansions in jobs. Blending highly sought after skill sets and capacity with the most recent innovations can help a region attract investment while strengthening the local economy. High-growth sectors will call for a high value staff. But work in these fields will be centered places of the United Kingdom. In 2027, the computer science and consulting sectors will account for just 9.9% of all job chances in the Northeast, 8.2% in Wales, as well as and 8.9% in the Republic of Ireland. Nevertheless, in the rest of the Southeast, they will account for 16% of the job prospects, and in London, they are expected to make up 23.1%. (Webber, 2024)

The NHS is running short of staff for about 125,572 in number. The existing ratio of patient to healthcare staff is 1 to 17 but in real the requirement is 1 to 11 which indicates more staff is to be required in the future. This requirement is increased considerably due to the covid-19 pandemic (Lucina Rolewicz, 2024).

The information seems to support higher spending in districts that have reduced employment growth and ensuing prosperity," stated Peter Arnold and the head of Economics at EY UK. Urging people to return to work could assist to certainly somewhat narrow the growth gap in some locations, however it's unclear which way the correlation lies and whether elevated inactivity causes decreased growth or the other way around. To address this, politicians are going to have to focus on a combination of health endeavors to address the substantial number of citizens in the UK who have left employment in recent years for reasons of permanent illness, along with programs and subsidies to increase membership (Webber, 2024).

The employment generated in London was far less than rest of the UK by 110,000. It has been projected by 2030, London will alone create 100,000 jobs due to high rate of migration. The number also increased after 2012. During the phase of 2013 to 2016, there was 0.9 % increase jobs growth in London because 580,000 more employment opportunities occurred (Neal Kilbane, 2017)

2.4 Conceptual framework

In this analysis, the conceptual framework revolves around taking certain factors into consideration for instance the economic indicators such as GDP growth, inflation, recession and pandemic. These factors played a massive role for the sudden change of the trends in the research which was the Recession of 2008, Pandemic and Brexit.

2.5 Gaps in existing literature

I have read and done the analysis of the existing literature which I found certain areas where more work needs to be done.
The existing literature mentioned in **2.3** with citations lack the following areas: -

- The data collected from surveys rather than from authentic government website.
- The machine learning models were not implemented to do get the data of forecasting for the years to come.
- Very few trends are studies, lack of data for the reason of sudden hike or downfall in trend across various sectors.
- There was no sign of using Power BI and Python to do forecasting.
- In the present literature, the connection between different sector employment was not provided.
- Most of the reports concentrated about a particular sector.
- Studies were conducted for London but its comparison with UK was very minimal.
- The figure facts were from open source as there was no reliable government source for them.
- Python time series ARIMA model was not implemented.
- The dataset I cleansed is more flexible, new data can be populated and would be automatically fetched by the libraries while implementing the machine learning model.

2.6 How my data analysis addresses the disparities of the current studies

In my data analysis and machine learning building, I implemented numerous tasks to make sure that the prediction for the data is accurate and reliable. The techniques and analysis I used are:

- Application of *Time series* model for the forecasting of the information
- Applying *Linear Regression* model for forecasting.
- Using *Python Script* in *Power BI* to visualize the data which generates the bigger picture.
- Creation of *ARIMA* model in *Python* for the forecasting of data.

- Studying the previous sudden uptrends and downtrend trends across all sectors and identify patterns and reasons.
- Comparing the *UK* and *London* data to find some valuable and common information.
- Cleansing and transformation the data for the purpose of creating accurate forecasting.
- In my research analysis, I also explained the code used for implementing the data.

CHAPTER 3: RESEARCH METHODOLOGY (DATA ANALYSIS LIFE CYCLE)

Data analysis life cycle has been used in this study which include sequence of steps to fetch, study and performing the analysis of the data. The primary objective of using this is to analyse the dataset of London and UK industry wise jobs to gather some valuable insights.

3.1 Discovery

This is first phase of the data analysis life cycle deals with the examination of the problem. It is extensively important to understand and know the data before performing analysis(geekforgeeks, 2022). The dataset, I chose is from Office for National Statistics. The data type for both the London (**Figure 3.1)** and UK dataset **(Figure 3.2)** is quantitative with number of jobs in thousands industry wise. The data is segregated as per quarterly basis from the year *1996* to ***2023***. The dataset of *London* and *UK* is available in *.xlsx* format(Labour Market Team, 2023).

There were certain prerequisites that had been kept in mind while selecting the dataset which was it should not contain empty values, collected from authentic source and must have more recent information.

	A	B	C	D	E	F	G	H	I
1	JOBS05 Workforce jobs [1] by region & industry								
2	Standard Industrial Classification (2007)								
3	Region - London								
4									
5			Agriculture, forestry & fishing	Mining & quarrying	Manufacturing	Electricity, gas, steam & air conditioning supply	Water supply, sewerage, waste & remediation activities	Construction	Wholesale & retail trade; repair of motor vehicles and motorcycles
6	SIC 2007 section		A	B	C	D	E	F	G
7	Mar 96 (r)		4	7	276	8	13	211	565
8	Jun 96 (r)		4	6	262	8	14	211	568
9	Sep 96 (r)		4	5	265	8	15	207	572
10	Dec 96 (r)		4	5	260	7	16	206	574

Figure 3.1 Dataset containing workforce jobs by region and industry in London on quarterly basis

	A	B	H	I	L	P	R	S
1	JOBS05 Workforce jobs [1] by region & industry							
2	Standard Industrial Classification (2007)							
3	Region - United Kingdom							
4								
5			Construction	Wholesale & retail trade; repair of motor vehicles and motorcycles	Information & communication	Administrative & support service activities	Education	Human health & social work activities
87	Mar 16 (r)		2,229	5,015	1,387	2,945	2,960	4,284
88	Jun 16 (r)		2,239	5,038	1,423	2,959	2,961	4,285
89	Sep 16 (r)		2,207	5,004	1,418	2,946	2,959	4,273
90	Dec 16 (r)		2,250	4,992	1,433	2,975	2,970	4,283

Figure 3.2 Dataset containing workforce jobs by region and industry in UK on quarterly basis

3.2 Data Preparation

Data preparation plays a very handy role to process the unprocessed or raw data. The data further transformed into information which enables business to identify upcoming trends, generate productivity. The procedure used for data preparation is *ETL (Extract, Transform & Load).* It is a method for data integration that merges the required information, cleans it and sorted the data from various sources(IBM, 2024).

- **Extract:** The unprocessed data is fetched from the source (Labour Market Team, 2023) into the staging area. After that the data is being sent for further processing
- **Transform:** In this phase, while being in the staging area the data is being cleansed by removing the unrequired information, unformed, validated and by modifying the column and the row names
- **Load:** The processed data is being transferred into the target area or software where set of modelling techniques will be implemented to get the prediction information and valuable insights from the data(IBM, 2024).

Data cleaning is an important aspect before processing the data. If the data is not processed and not cleaned, then the forecasting outcome would not be accurate. It's waste in, waste out in the field of data science and analytics. This suggests that the quality of the output depend on the quality of the data eating, irrespective of the sophisticated our analytics or forecasting algorithms are. Every one of the above processes rely on data, so it's vital to take the time necessary to ensure that the data is well and used for functionality. Putting it plainly, data cleaning, commonly referred to as cleansing, is a step essential before data analysis is allowed to begin. This may entail finding and removing empty and duplicate entries along with modifying data to repair false records. Although an issue of tainted or illicit information has been with us through the past ten years, the generation of data has increased tremendously. worldwide produced over 64 zettabytes of data in 2020 alone, or 6.4 trillion gigabytes. Even if fewer than 2% of all this data is kept, it undoubtedly aggravates the generally accepted issue of "dirty data" worse. It's vital to comprehend how information got infected in the first place before undertaking any extensive cleaning. This might come from ordinary sources. Cleaning up information is an essential task that aids in developing of a template for data cleansing inside the company. Any statistical analysis, information science process is trash in, trash out, as was previously stated. disregarding it leads to expensive, faulty analytical outcomes that need invested time, money, and additional assistance.

To make the data clean for better assessment following steps have been taken:

- **Eliminating non required regions**: The dataset workbook contains the information about 15 regions on separate worksheets out of which only the London and UK worksheet is kept and other was removed. The worksheet name from *7. London* has been changed to *london_jobs_data and 15.* United Kingdom *to uk_jobs_data* **(Figure 3.3)**

Figure 3.3 Worksheets name updated

- **Taking out top rows:** The top 4 rows has been removed so that the only the column names becomes the headers or title rows for the data set**(Figure 3.4)**

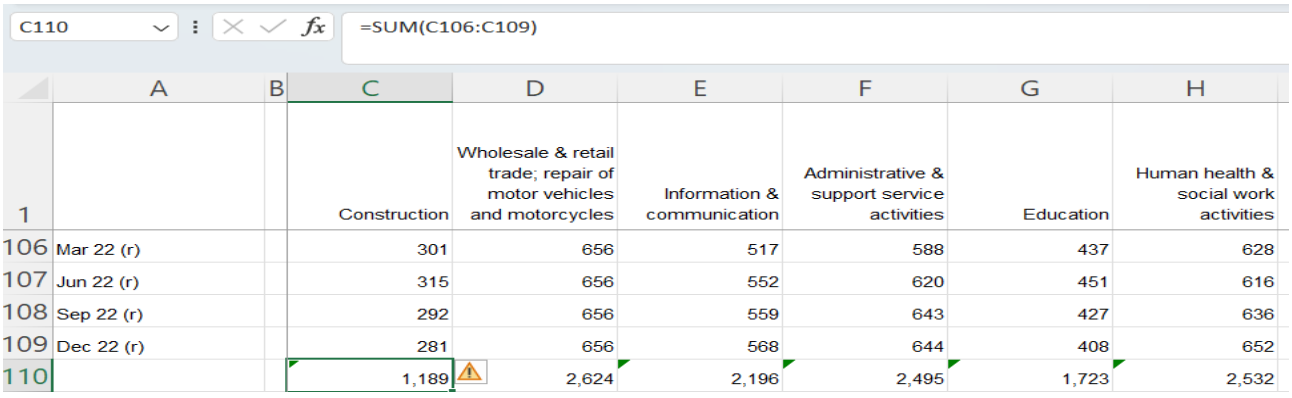
Figure 3.4 Top fours removed from London and UK worksheet

- **Keeping selected industry sectors:** The dataset comprises the information about 20 industry sectors for both London and UK worksheet from which only 6 are kept (**Figure 3.5)** and others have been removed. The reason for doing this to visualize the data correctly which looks clean, tidy and uncluttered. The industry sector information selected for data processing are *Construction, Wholesale & retail trade - repair of motor vehicles and motorcycles, Information & communication, Administrative & support service activities, Education*and*Human health & social work activities.*

B	C	D	E	F	G
Construction	**Wholesale & retail trade; repair of motor vehicles and motorcycles**	**Information & communication**	**Administrative & support service activities**	**Education**	**Human health & social work activities**

Figure 3.5 Industry sector selected for data analysis of London and UK

- **Converting quarterly data to yearly:**To have better analysing of data, the four quarters of a year have been added to transform it into the year data for instance the all the 4 quarters March, June, September and December of the year 2022 have been added using excel formula *=SUM(C106: C109)* in the cell address *C110* and drag the result to right side for all the industry sector year values **(Figure 3.6).** This has been implemented for both the London and UK dataset.



	A	B	C	D	E	F	G	H
1			Construction	Wholesale & retail trade; repair of motor vehicles and motorcycles	Information & communication	Administrative & support service activities	Education	Human health & social work activities
106	Mar 22 (r)		301	656	517	588	437	628
107	Jun 22 (r)		315	656	552	620	451	616
108	Sep 22 (r)		292	656	559	643	427	636
109	Dec 22 (r)		281	656	568	644	408	652
110			1,189	2,624	2,196	2,495	1,723	2,532

Figure 3.6 Quarterly data added and transformed to year

- **Detecting and handling the missing values**: The dataset does not have the values for the last quarter for the year 2023. Therefore, the incomplete quarterly data has been removed from the both the London and UK data (**Figure 3.7)** to make it more precise and uniform.

	A	B	C	D	E	F	G	H
1			Construction	Wholesale & retail trade; repair of motor vehicles and motorcycles	Information & communication	Administrative & support service activities	Education	Human health & social work activities
111	Mar 23 (r)		284	656	570	674	430	692
112	Jun 23 (r)		281	656	571	681	416	689
113	**Sep 23 (p)**		**279**	656	**578**	**679**	**418**	**698**

Figure 3.7 Quarterly data for 2023 removed for not having December information

- **Rename the column to Year:** After adding all quarters, the column name is set to *Year* and all the values generated from formula must be copied and pasted in the same cell as *Paste as Values (***Figure 3.8)**. The reason for doing that is because if the quarters are removed then the value generated from formula becomes 0.

	A	B	C	D	E	F	G
1	**Year**	**Construction**	**Wholesale & retail trade; repair of motor vehicles and motorcycles**	**Information & communication**	**Administrative & support service activities**	**Education**	**Human health & social work activities**
2	1996	835	2,279	986	1,473	914	1,415
3	1997	806	2,387	1,096	1,568	959	1,426
4	1998	909	2,536	1,183	1,706	994	1,380
5	1999	879	2,659	1,314	1,799	1,073	1,353
6	2000	958	2,719	1,404	1,929	1,092	1,413
7	2001	950	2,673	1,452	1,971	1,054	1,480
8	2002	938	2,610	1,390	1,881	1,120	1,478
9	2003	948	2,561	1,384	1,843	1,173	1,546
10	2004	961	2,549	1,288	1,851	1,222	1,600

Figure 3.8 Yearly data for final analysis

3.3 Plan Model

In this phase, the study of variables has been performed for the purpose of identifying the relationship between them. Apart from that dependent variable and independent variable is also figured out(geekforgeeks, 2022). The methods which I have used for the analysis the data set are ***Time series*** and ***Regression***.

The ***Time series*** is an array of data points that occur one after one other over an interval of time. Data from cross-sections, on the reverse hand, reflects a single point in time. Time series is a handy data analysis technique which has been applied on the data set to analyse how the number of jobs in particular industry changes over time. The time series analysis works better when the data is arranged in a particular set of intervals (HAYES, 2022).

On the other hand, ***Regression*** is a statistical data analysis technique which is used understand the relationship between the ***dependent variable*** and the ***independent variable***. In the dataset, the number of jobs in sector is considered as a dependent

variable because these are the variables that needs to be forecasted whereas the column which contains quarterly data is considered as an independent variable. ***Slope*** is used to see the impact of dependent variable when the value in independent variable changes. The ***Y-intercept*** assess the value of one variable when the other variable value is zero (BEERS, 2024).

3.4 Model Building

Model 1 - Regression

A connection between a few independent variables and the reaction, dependent, or target variable can be defined by a function that can be provided by a model known as regression. As an example, an equation based on linear regression could be used to unravel the link between weight and height. Regression modelling serves as the basis for a variety of methods for predicting and the examination of effects of target variables. Regression models are frequently used to support studies reported in news outlets about topics such as gas mileage, the cause of contaminants, and the effect of multimedia consumption on cognition. A theory with a straight line as the link between the inputs and the outcomes is known as a linear regression. This is quite straightforward for comprehending and even witness in practice. Our brains tend to identify trends in associations even when they lack particularly linear and apply an essential linear model to everything.

The concept of multiple regression findings indicate that every input variable may have an influence on the target variable, or output. You may add an attribute with the total quantity of emails sent in the previous month for our email advertising example. Evaluating both input variables help to paint a more thorough understanding of what motivates clients to reply to a campaign and how best to timetable and send out emails. Even though its conceptualization grows more complicated with additional inputs, a linear relationship may remain. Learning the precise effects of every parameter along with the way they combined to produce the desired end variable results is crucial for these types of models (IMSL, 2021).

The first model I used is *Regression* which is a statistical and data analysis method to find the relationship between the *dependent variables* and an *independent variable*. It not only finds the relationship but also find how depth the relationship between dependent and independent variable is.

Dependent variable is the one whose outcome is being measured. The value of dependent variable is totally relying on the independent variable. As far as the regression model equation is concerned, its value is always represented on ***y-axis***. In the data set of ***London*** and ***UK***, the dependent variables are different industries of job sector which are ***Construction, Wholesale & retail trade repair of motor vehicles and motorcycles, Information & communication, Administrative & support service activities, Education and Human health & social work activities***.

Independent variable is the one which creates an impact on dependent variable. In the regression model equation, independent variables are always represented on ***x-axis***. In the dataset, the independent variable is the ***Year.***

Several applications, connectors, and services called Power BI come combine for transforming the different sources of data into integrated, interactive, and visually appealing insights. Your data may be maintained in a hybrid data warehouse that runs premises in the cloud, or it could be in be stored in an Excel document. With Power BI, business can view analyse what's essential, integrate to your data streams with at ease and share this data with whomever you select. Generating a report and attaching to data sources in Power BI Desktop are the first phases in a typical Power BI operation. With that, you send and upload the graph from

Power BI Desktop to the Power BI service, permitting customers to examine and interact with it on portable devices as well as in the Power BI service. It can be used to view the reports and dashboards (Microsoft Build, 2024) .

For model building I used *Power BI.*

Steps performed for implementing the *Regression* in ***Power BI*** are: -

1. Launch the ***Power BI*** desktop.
2. Upload the data set ***London_jobs_data*** and ***uk_jobs_data*** from ***Get data*** option.
3. Now select the ***Scatter chart*** from the ***Visualisation pane*** (**Figure 3.9)**

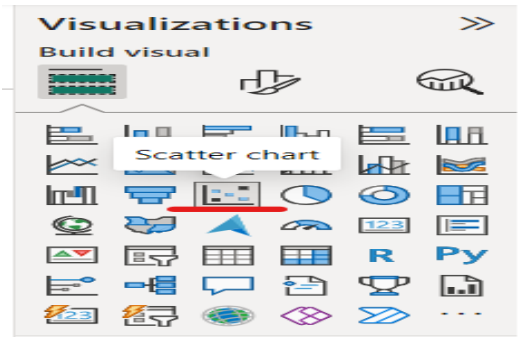


Figure 3.9 Scatter chart icon in visualisation

4. Now put ***Year*** which is ***independent variable*** on ***x-axis*** and ***Construction*** column which is ***dependent variable*** on ***y-axis*** (**Figure 3.10)**

Figure 3.10 Setting dependent variable and independent variable on their axis

5. The trend line setting **(Figure 3.11)** and Scatter chart with trend line **(Figure 3.12)** is to be attached which shows the overall trend for the number of jobs in construction sector generated each year. This can be done by going to ***Visualisation pane*** – ***Analytics Tab*** - ***Enable*** the ***Trend line*** option.

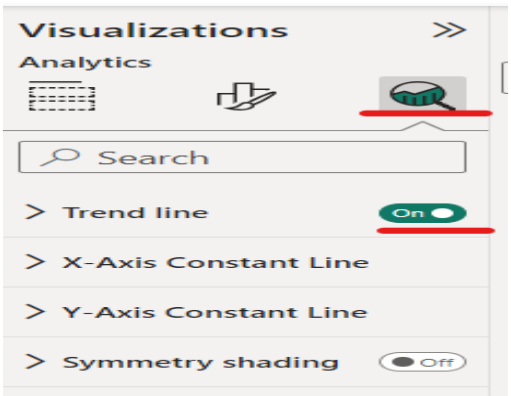


Figure 3.11 Trend line option in Analytics pane

6. The trend line on Scatter chart (**Figure 3.12)**

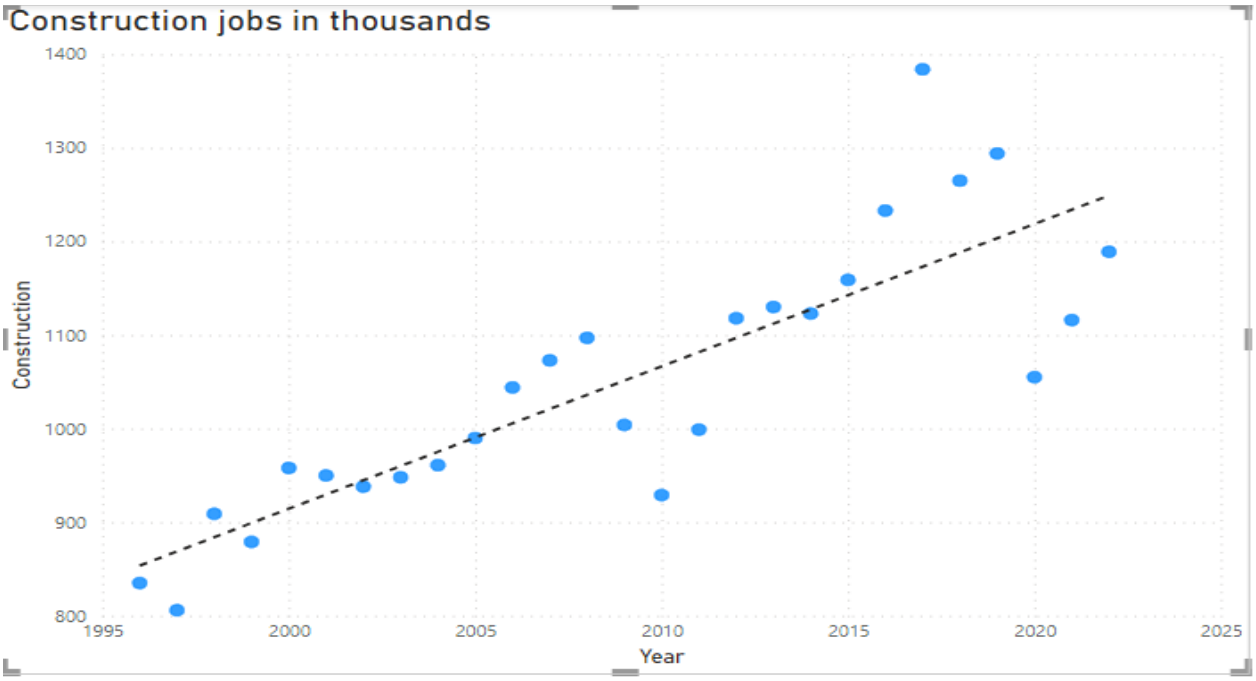


Figure 3.12 Trend line displays the trend

7. In the next step, the ***Slope*** and ***Intercept*** are generated using ***LINESTX*** function. This can be done by creating a new table named ***info*** from the Modelling tab and write the code snippet (**Figure 3.13)**

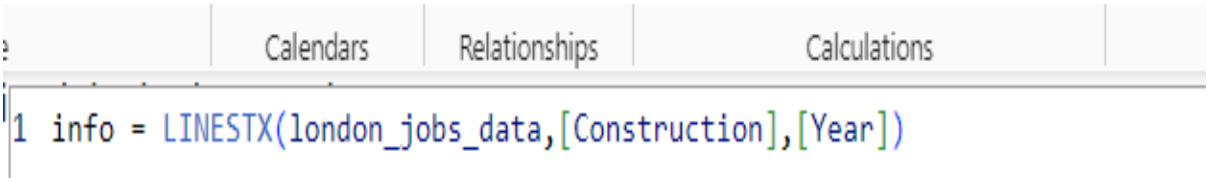


Figure 3.13 Code for implementing LINESTX function

Code description:

Info is the name of the table, ***LINESTX*** function contains 3 arguments: London_jobs_***data*** as a table name, ***[Construction]*** acts as a variable for ***y-axis*** and ***[Year]*** act as a variable for ***x-axis.***

8. The code snippet generates the info table with various values (**Figure 3.14)** by clicking on the ***Table view.***

Slope1	Intercept	StandardErrorSlope1	StandardErrorIntercept	CoefficientOfDetermination	StandardError	FStatistic	DegreesOfFreedom	RegressionSumOfSquares
15.2136752136752	-29512.9401709402	1.93620611166343	3889.8673124606	0.711781621285566	78.3625455623319	61.7397842965799	25	379124.786324

Figure 3.14 The table generated columns and values for slope and intercept

9. Now it's time to create a ***new parameter*** which enables the user to drag the slide to the particular year and see the prediction number in the corresponding card. I set its name as ***Select_Year***. The values are set as per (**Figure 3.15)**

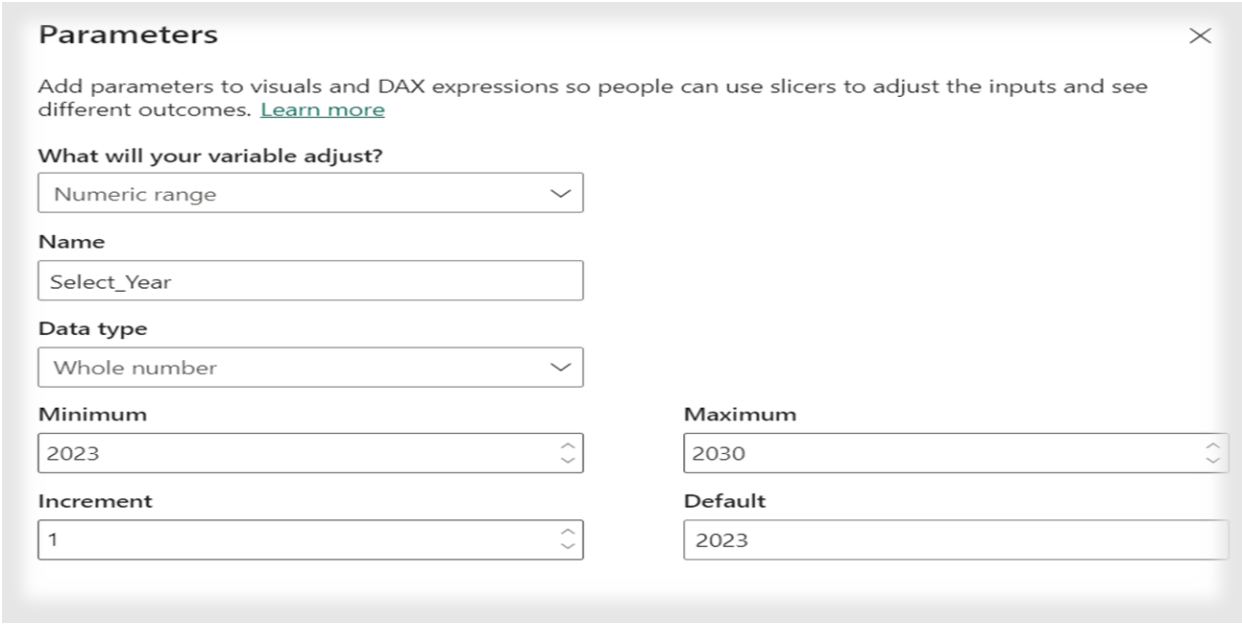


Figure 3.15 Values set for the parameter

10. As a result, it is going to create a Slicer (**Figure 3.16)**

Figure 3.16 Slicer with default value 2023

11. Equation to use the regression (**Figure 3.17)** to forecast the number of jobs is created.

```
1 regression construction =
2 VAR _linest = LINESTX(london_jobs_data, [Construction], [Year])
3 VAR _slope = SELECTCOLUMNS(_linest,[Slope1])
4 VAR _intercept = SELECTCOLUMNS(_linest, [Intercept])
5 RETURN
6 _intercept + _slope * 'Select_Year'[Select_Year Value]
```

Figure 3.17 Regression equation code

Code description:

The above codes predict the number of jobs for the year 2025 in construction sector. The 3 variables ***_linest, _slope*** and ***_intercept***are created which fetch the value of ***Slope1*** and ***Intercept.***

The value for Return is line equation which is ***m * x + b*** where ***m*** is ***slope***, x is independent variable ***Year*** and b is ***intercept.***

12. Drag the card from the Visualisation pane and then drag the equation to the card to see the numerical value for the number of jobs in construction. The predicted value shown on the card depends upon the ***Year*** selected in the ***Slicer*** for the year ***2024*** and ***2025*** (**Figure 3.18)**

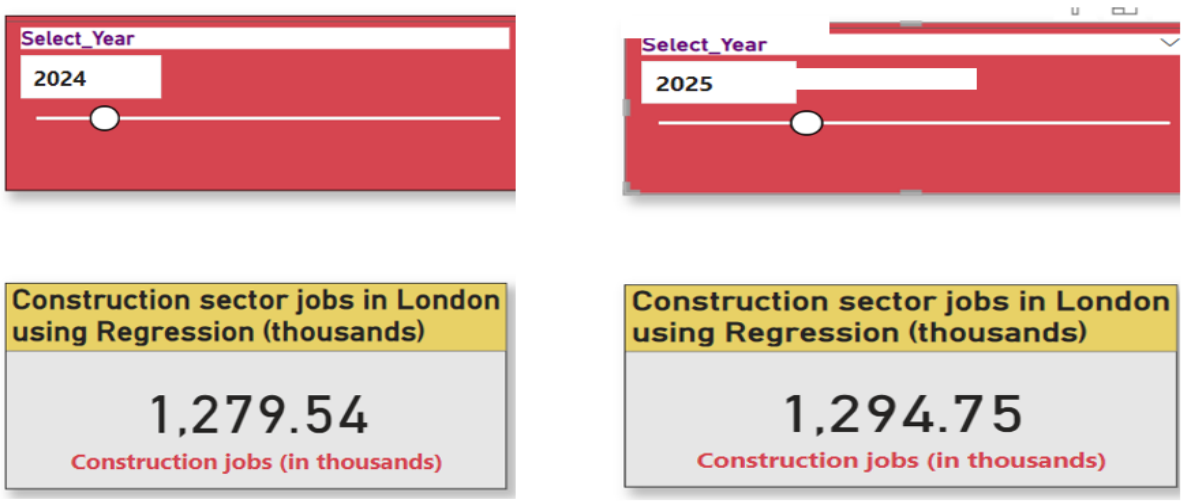


Figure 3.18 Jobs prediction in year 2024 and 2025 by changing slicer value for year

Regression Analysis for predicting the number of jobs in Construction sector in UK

1. Set the values of ***uk_jobs_data*** data set in scatter plot as ***Year*** on ***x-axis*** and ***Construction*** jobs on ***y-axis***.Enable the trend line**(Figure 3.19)**

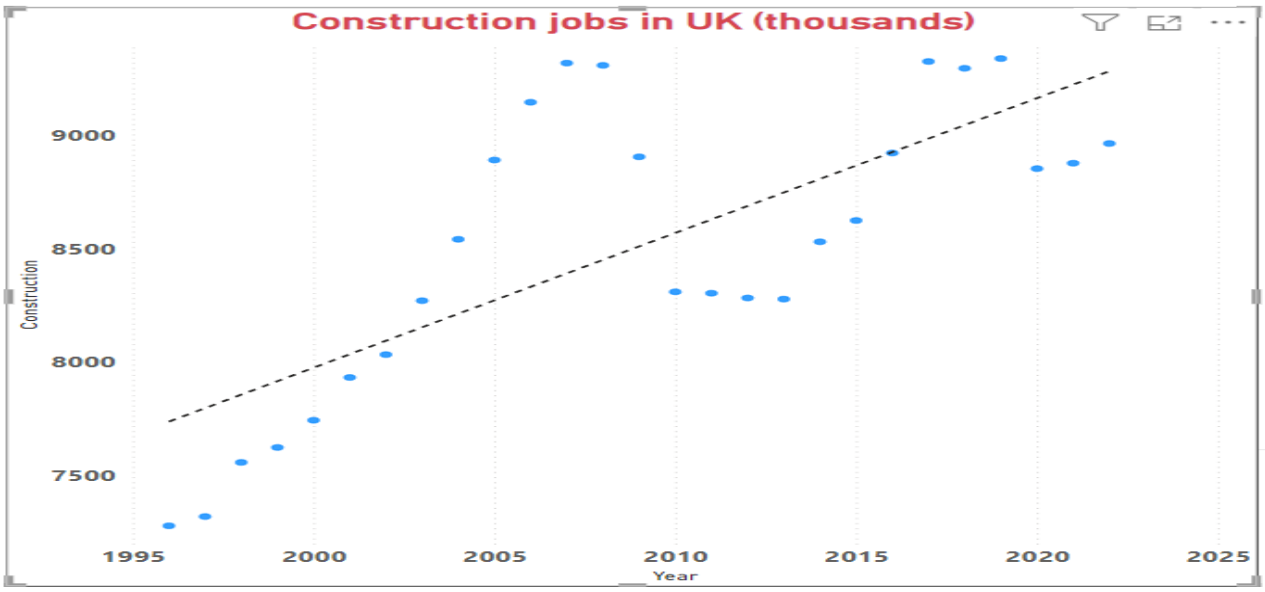


Figure 3.19 Trend line shows the dataset trend

2. Make a table for getting the Slope and Intercept value (**Figure 3.20)**

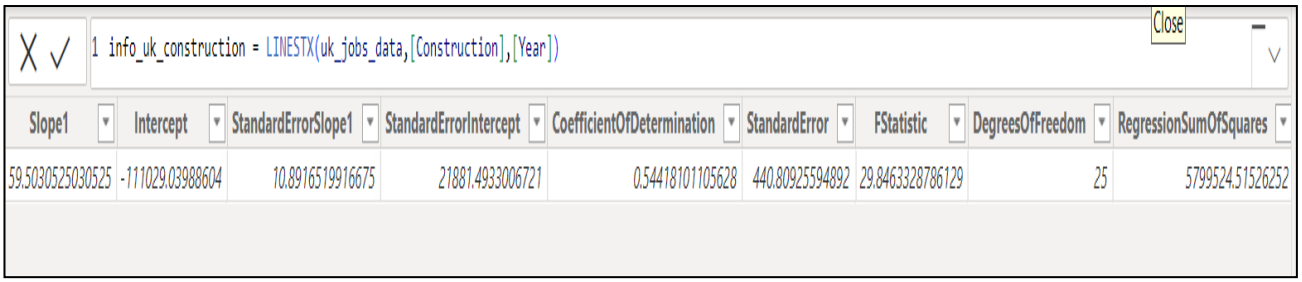
```
1 info_uk_construction = LINESTX(uk_jobs_data,[Construction],[Year])
```

Figure 3.20 LINESTX function on uk_jobs_data

Code description:

A table named ***info_uk_construction*** is created in which I used the ***LINESTX*** function which take 3 arguments which has the dataset name ***uk_jobs_data***, ***Construction*** column on ***y-axis*** and ***Year*** column on ***x-axis***.

3. The code snippet (**Figure 21) generatesthe** table with the values (**Figure 3.21)**



Slope1	Intercept	StandardErrorSlope1	StandardErrorIntercept	CoefficientOfDetermination	StandardError	FStatistic	DegreesOfFreedom	RegressionSumOfSquares
59.5030525030525	-111029.03988604	10.89165199166675	21881.4933006721	0.544181011056628	440.80925594892	29.8463328786129	25	5799524.51526252

Figure 3.21 Slope and Intercept value generated by LINESTX function

4. Equation to use the regression to forecast the number of jobs in construction sector (**Figure 3.22)**

```
1 regression_construction_uk =
2 VAR _linest=LINESTX(uk_jobs_data,[Construction],[Year])
3 VAR slope=SELECTCOLUMNS(_linest,[Slope1])
4 VAR intercept=SELECTCOLUMNS(_linest,[Intercept])
5 RETURN
6 intercept + slope*Select_Year_UK[Select_Year_UK Value]
```

Figure 3.22 Code snippet returning Regression equation line

Code description:

The above codes predict the number of jobs for the year 2025 in construction sector. The 3 variables ***_linest, _slope*** and ***_intercept***are created which fetch the value of ***Slope1*** and ***Intercept.***

The value for Return is line equation which is ***m * x + b*** where ***m*** is ***slope***, x is independent variable ***Year*** and b is ***intercept.***

5. Drag the card from the Visualisation pane and then drag the equation to the card to see the numerical value for the number of jobs in construction. The predicted value shown on the card depends upon the ***Year*** selected in the ***Slicer*** for the year ***2024*** and ***2025 (*****Figure 3.23)**

Figure 3.23 Changing value in slicer shows predicted values for the year 2024 and 2025

Model -2 Time series

The steps to perform time series analysis for predicting the number of jobs in Construction are :-

1. Go to ***Visualisation pane*** and select ***Line chart*** as shown (**Figure 3.24)**

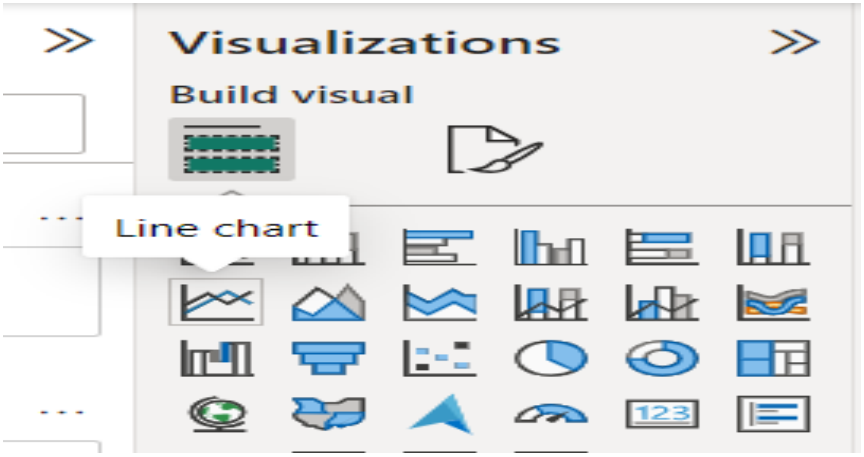


Figure 3.24 Line chart icon in Visualisation pane

2. Then select the ***london_jobs_data*** table and drag ***Year*** on ***x-axis*** and ***Construction*** column on ***y-axis (*Figure 3.25)**

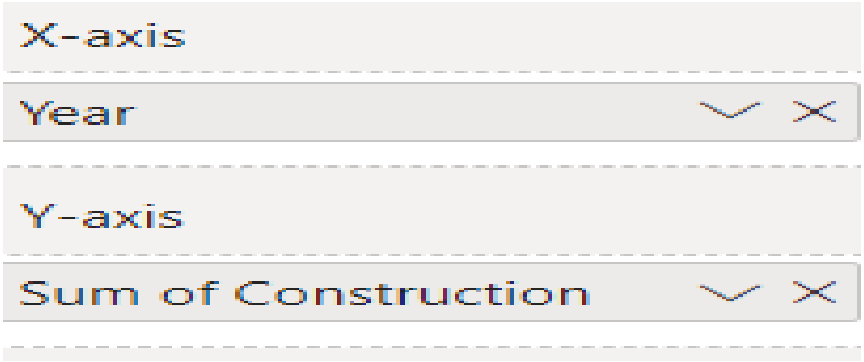


Figure 3.25 Setting value for x axis and y axis

3. Then the ***Trend line*** is ***enabled*** from ***Analytics tab.***
4. Go to ***Analytics tab*** and enable the ***Forecast*** option.
5. The ***Time series*** analysis predicts the number of jobs in construction in year ***2024*** as shown in following (**Figure 3.26)**.

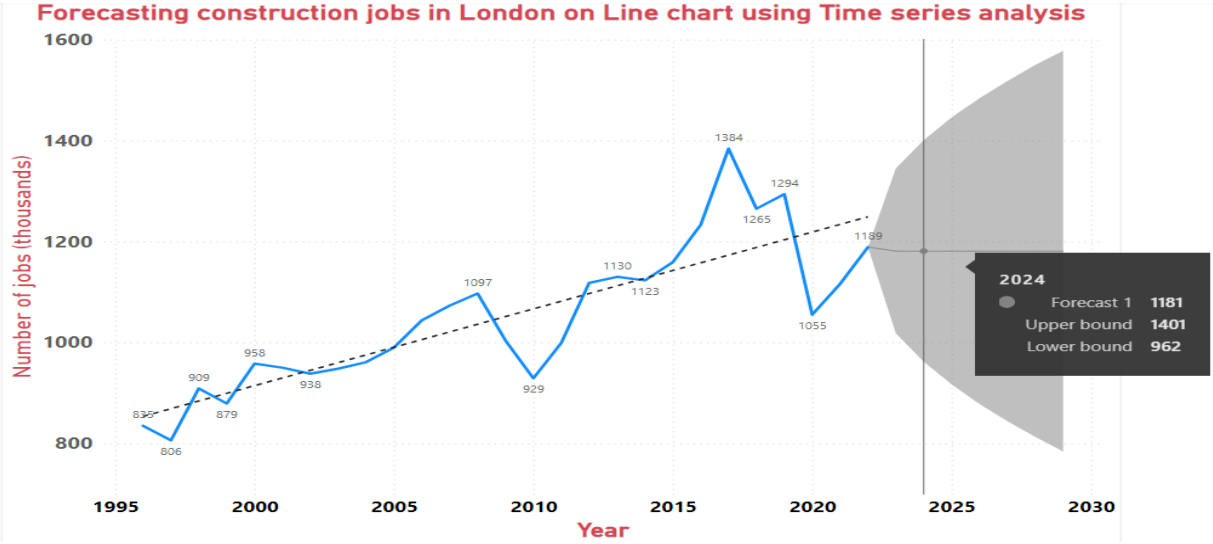


Figure 3.26 Forecasting value for year 2024 appeared within gray area for London

Forecasting the construction jobs in UK

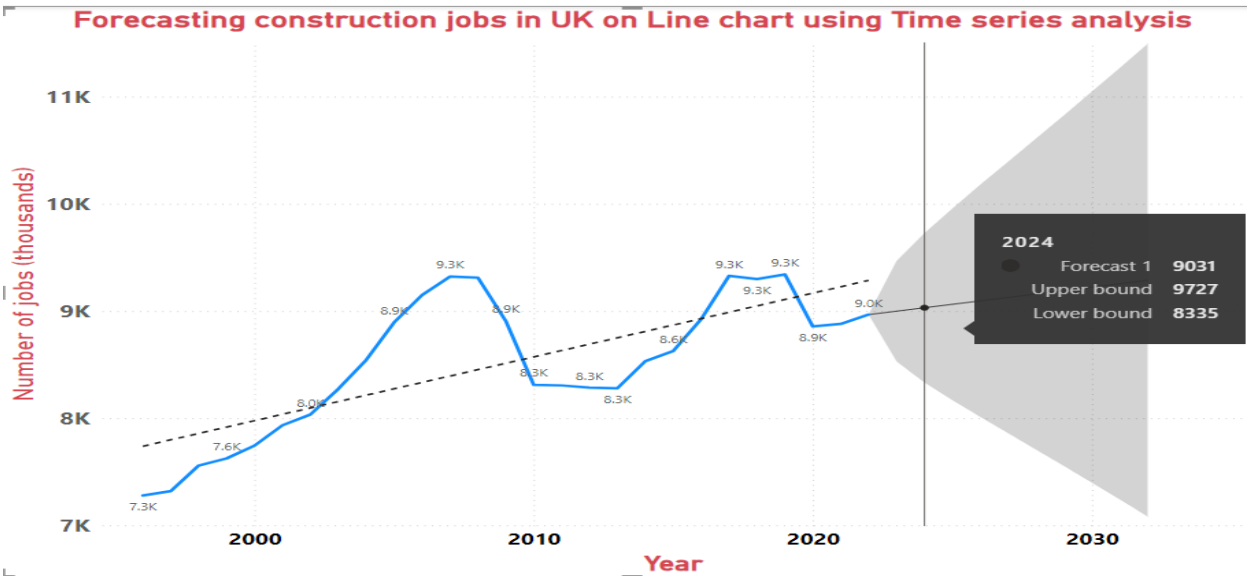


Figure 3.27 Forecasting value for year 2024 appeared within gray area for UK

3.5 Communication results

In this step, the resultant information from different models is fetched and compared. Based on comparison and analysing the type of data, the best model is to be selected. The best model will be providing some valuable information to the concerned authorities(geekforgeeks, 2022).

The data is being analysed by ***Time series*** and ***Regression*** modelling techniques. As far as the model selection is concerned, the *time series* model has been given more priority than the Regression model. The foremost reason for choosing *time series* is that both datasets have an attribute ***Year*** which is entirely dependent on time. This further enables the data scientist to generate the

prediction pattern based on different year values keeping in track of all previous sudden high and low trends. Eventually, it is going to provide the predictions which is more likely to be accurate.

On the other hand, ***Regression model*** is entirely focusses on relationship of dependent variable with the independent variable. The reason for not choosing Regression model is that the independent variable Year keeps on constantly increasing by one which doesn't provide any accurate impact on the dependent variable in which the number of jobs is predicted. The comparison between the predicted values on London dataset has been done by using Regression and Time series analysis (**Figure 28)** as it can be seen in Regression the prediction number increases at the constant rate of approximately 15 thousand as the year progresses which is not the appropriate scenario.

Year	Regression	Time series
2024	1279.54	1401
2025	1294.75	1446
2026	1309.97	1484
2027	1325.18	1518
2028	1340.49	1550

Figure 3.28 Construction jobs (thousands) in London

This is an important aspect of the data analysis. The data is being gathered and eventually it is going to generate plenty of insights. The insights and prediction number therefore going to help the government officials, policy makers, individual job seekers. The table (**Figure 3.29)** shows the number of jobs (in thousands) predicted across various sectors in ***London usingTime Series Analysis***for the year range from ***2024*** to ***2028***.

Year	Construction	Wholesale & retail trade; repair of motor vehicles and motorcycles	Information & communication	Administrative & support service activities	Education	Human health & social work activities
2024	1401	2862	2310	2532	1790	2637
2025	1446	2906	2367	2563	1823	2687
2026	1484	2943	2424	2594	1856	2737
2027	1518	2976	2481	2625	1889	2787
2028	1550	3007	2538	2656	1922	2837

Figure 3.29 Predicted number of jobs in thousands in London

The table (**Figure 3.30)** shows the number of jobs (in thousands) predicted across various sectors in ***UK*** using ***Time Series Analysis*** for the year range from ***2024*** to ***2028***.

Year	Construction	Wholesale & retail trade; repair of motor vehicles and motorcycles	Information & communication	Administrative & support service activities	Education	Human health & social work activities
2024	9031	19720	6826	13124	12336	18985
2025	9063	19858	7106	13529	12454	19246
2026	9095	19975	7206	13431	12573	19507
2027	9127	20079	7396	13170	12691	19768
2028	9159	20174	7586	13434	12810	20029

Figure 3.30 Predicted number of jobs in thousands in UK

3.6 Operationalize

This is also called deployment. In this phase, the advantages of results being communicated to the higher authorities. The pilot project is ready to set for the implementation. The reports and implementation code are being delivered to the concerned department (geekforgeeks, 2022).

CHAPTER 4: BUILDING MACHINE LEARNING MODEL (PYTHON SCRIPT IN POWER BI)

4.1 About Python

Python is extensively used programming language and is used for the development of applications in the areas of Artificial intelligence, Machine learning. It is also considered as high-level language and was developed by *Guido Van Rossum* in the tenure of *1985 - 1990.* Python plays a massive role in data science for the purpose of predicting and doing classification by implementing machine learning models. It is also open source which means its source code is available under GPL (General Public License). Big IT giant companies like IBM, Intel, Amazon, Netflix, Uber, PayPal, Pinterest, Nasa etc are using python in their development environment (tutorialspoint, 2022).

4.2 Python Merits

The primary advantage of python is that it is quite simple to learn, and its syntax is not even complex. It has plenty of advantages which are: -

- The scripts in python does not require to be compiled before execution and called as interpreted language.
- Python programs does not need to be compiled before execution.
- It supports object-oriented programming which allows creation of objects.
- This language has the super advantage of being portable which means a program made in Windows environment can successfully be transferred to LINUX environment without any hassle.
- Python has large set of inbuilt libraries such as pandas, matplotlib which has been used for the using analytical tasks and plotting various graphs to give the visual representation of the data (tutorialspoint, 2022).

4.3 Connecting Python with Power BI desktop

To run Python programs and scripts, it is the pre requirement to integrate Python with Power BI desktop. The steps for its implementations are as follows: -

1. Firstly, I installed Python in my system with version *Python 3.12.3* by creating a folder named *Python* in *C* drive and sets this as the installation path.
2. In the next stop, I installed libraries like *matplotlib* and *pandas*. This can be done by opening *Command Prompt* and then set the location as *cd Python\Scripts* (**Figure 4.1)** because the libraries are always installed in Scripts folder.

Figure 4.1 Setting location for installing libraries

3. The pip command is used followed by install and then the name of library. Here I am installing *pandas* library **(Figure 4.2)** which is extensively used for the manipulation and analysis of the data and after applying the command, the library is simply installed **(Figure 4.3)**

Figure 4.2 Installing pandas library

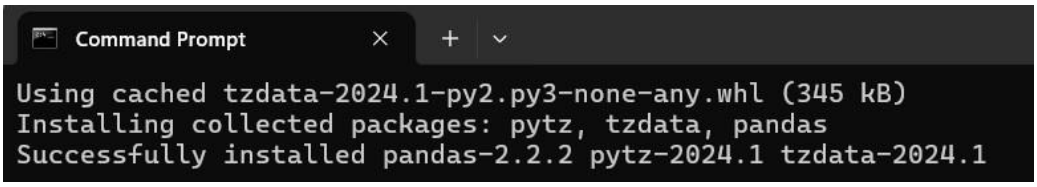


Figure 4.3 Prompt showing pandas library successfully installed

4. The another library which needs to installed in *matplotlib* which is used for the plotting of graphs and charts. This library is going to be installed by using command *pip install matplotlib* (**Figure 4.4)** and is successfully installed. **(Figure 4.5)**

Figure 4.4 Installation of matplotlib library

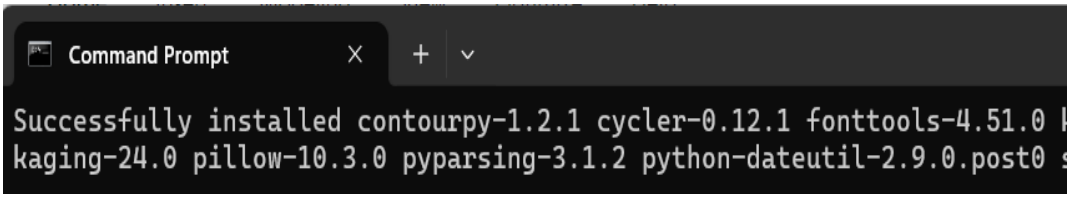


Figure 4.5 Command prompt displaying successful installation of matplotlib library

5. For loading the dataset which is *excel* format, there is need to install one more library which is pandas openpyxl (**Figure 4.6)** which enables to load, read the excel files and is gets implemented in python (**Figure 4.7)**

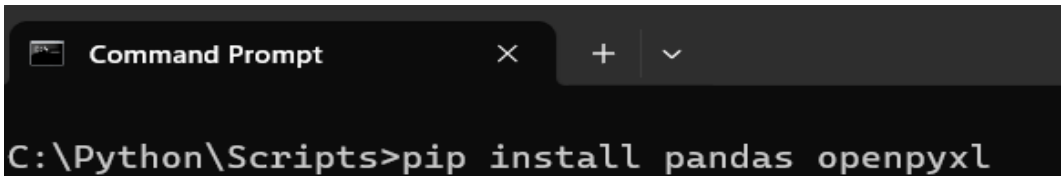


Figure 4.6 Command to install openpyxl library

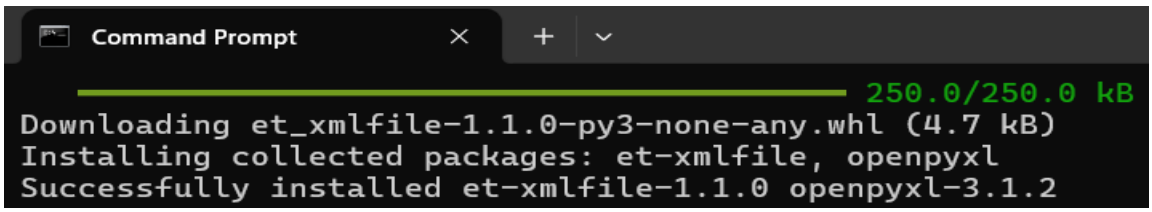


Figure 4.7 Successful installation of openpyxl library

6. In the next step, the *Power BI* is to be open. The new report is to be saved with the name *of python machine learning model*

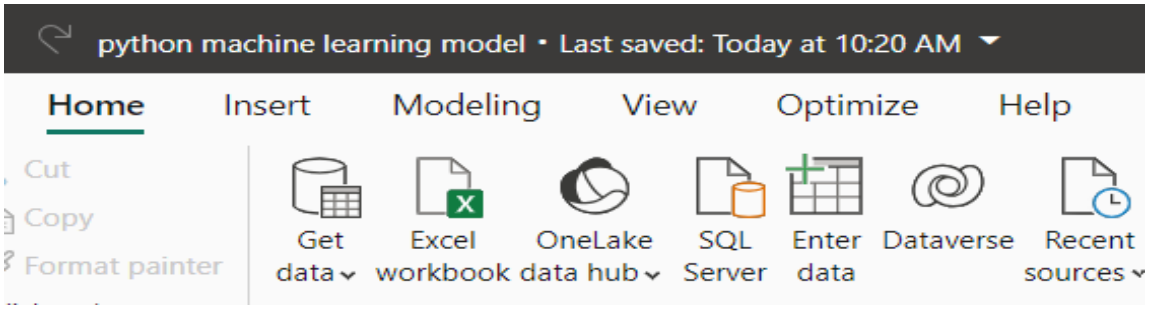


Figure 4.8 Creating and saving a new report

7. Then the installation location for the Python is to be set in *Power BI* which is C:\Python (**Figure 4.9)**

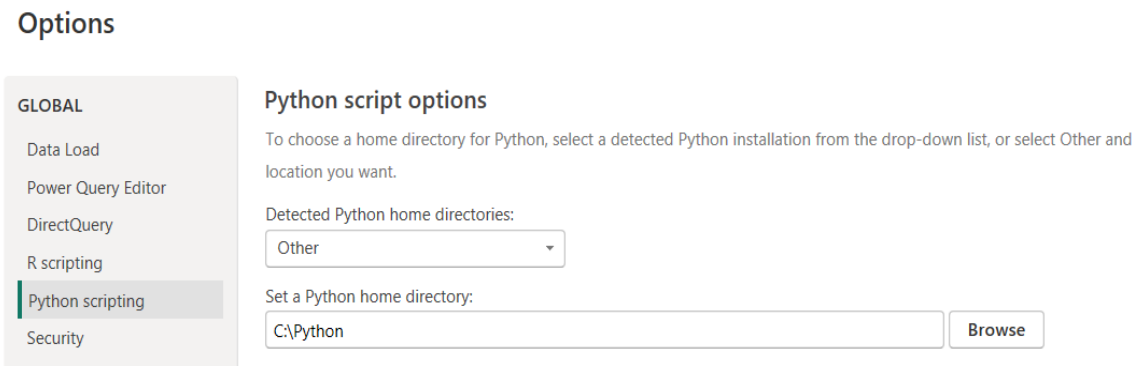


Figure 4.9 Path setting for Python

8. In this step, the Python is to be set a data source by choosing *Get Data* and search Python and click on *Python Script* and then *Connect* (**Figure 4.10)**

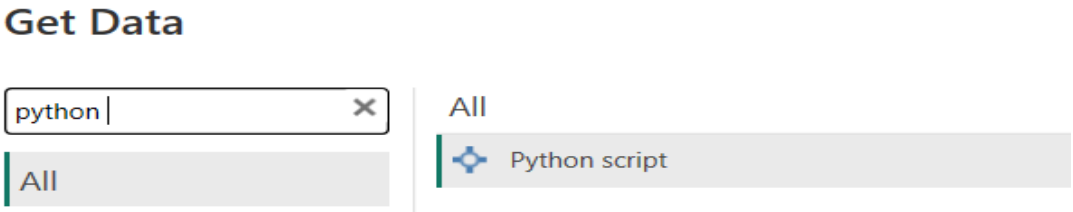


Figure 4.10 Choosing Python as a data source and establishing connection

9. The Python Script window will be open where the python code has been written where the dataset is going to be loaded in Power BI (**Figure 4.11**).

Figure 4.11 Python code to load the dataset in Power BI

Python script description (Figure 4.11)

- *import pandas as pd*: It is importing the pandas library.
- The variable *dataset* is created which is storing the path of the dataset *jobs05dec2023.xlsx*.
- Since the dataset *jobs05dec2023.xlsx* has 2 sheets which needs to be stored separately.
- The variable *sheet_london* is storing the data of *london_jobs_data*.
  The variable *sheet_uk* is storing the data of *uk_jobs_data*.
- The variable *london_dataset* needs to be created which is going to store the data which has been fetched by using function *pd.read_exce( )l* function which is going to have 2 arguments.
  The first argument is *dataset* which is a variable that stores the entire dataset and the second argument is the sheet_name which is *sheet_london.*
- The variable *uk_dataset* needs to be created which is going to store the data which has been fetched by using function *pd.read_exce( )l* function which is going to have 2 arguments.
  The first argument is *dataset* which is a variable that stores the entire dataset and the second argument is the sheet_name which is *sheet_uk*
- Then the variable containing the different sheets are printed by using command just writing the name of the variables *london_dataset* and *uk_dataset*

10. In the next step script is executed by pressing the *OK* button. It then opens the *Navigator* window which show the dataset name which were being fetched by using Python script **(Figure 4.12)**.

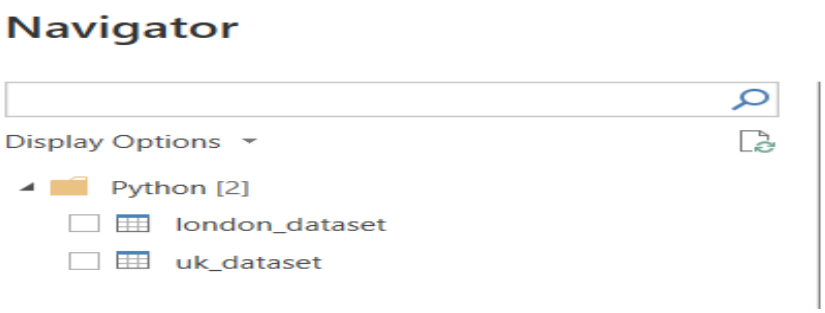


Figure 4.12 Navigator window shows the name of columns fetched

11. The datasets *london_dataset* and *uk_dataset* is to be selected and the preview of the data can be seen on the right side of the Navigator window **(Figure 4.13)**.

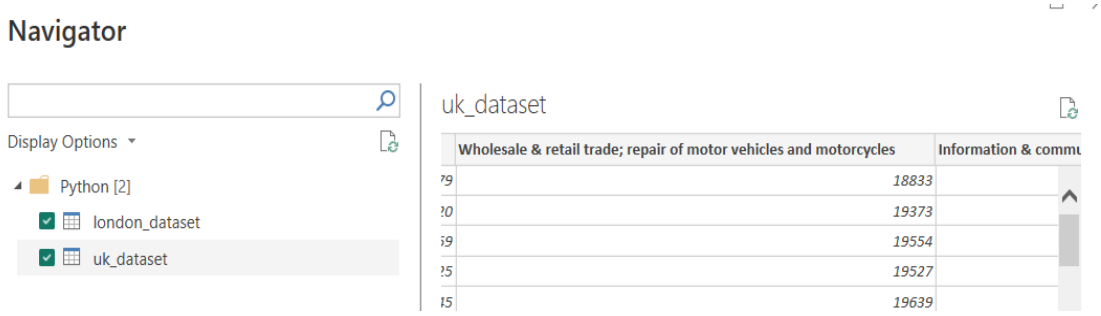


Figure 4.13 Preview for the datasets

12. Click on the Load button and both the datasets are loaded (**Figure 4.14)**

Figure 4.14 Datasets loaded in the Power BI

4.4 Creating visuals in Power BI from the data fetched by Python script

The data in table form do not convey the summary information. Therefore, it is a good practice to create the visuals to represent the data. The steps to create visuals are as follows: -

1. Go to *Visualisation Pane* and click on *Table* (**Figure 4.15)**

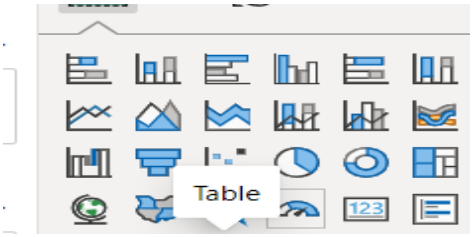


Figure 4.15 Table option in Visualisation

2. In the *london_dataset*, select the column *Education* and *Year.*

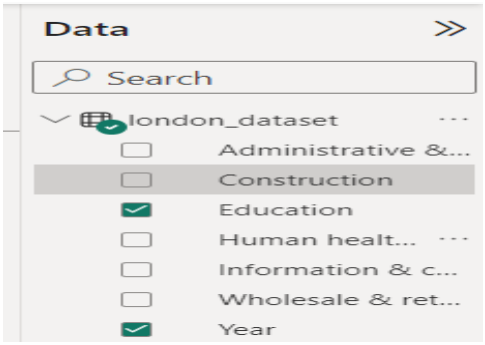


Figure 4.16 Select 2 columns.

3. The information fetched from columns is displayed in the table (**Figure 4.17)**.

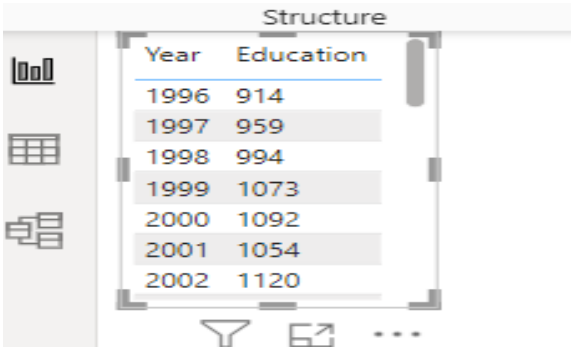


Figure 4.17 Table containing the data

4. In the *Visualisation* pane, click on Python visual *Py* icon.

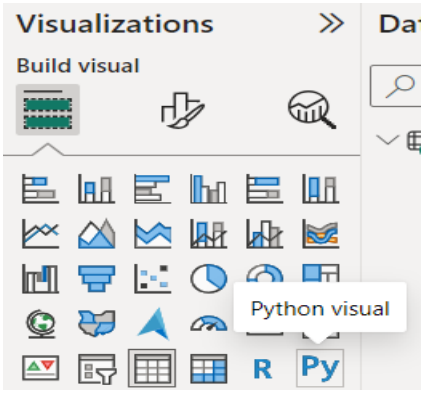


Figure 4.18 Python Visual icon

5. A python script editor opens and need to insert the code snippet to create a Scatter chart (**Figure 4.19)**.

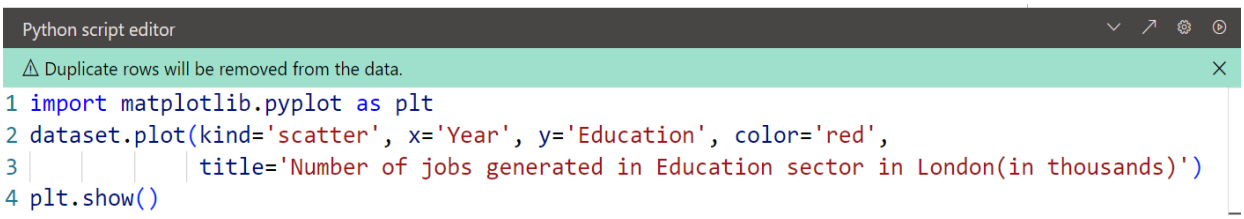


Figure 4.19 Python coding for creating a Scatter chart

Code description:

- There is a library which *is matplotlib.pyplot* is to be imported and then its object *plt* is to be created.
- The dataset is incorporated with plot function which takes 4 arguments *kind* stores the type of chart that needs to be created for instance here I created *Scatter* chart, *x* is the axis which contains the value of *Year* column and *y* is the axis which contains the value of the *Education* Column, the *color* is set to *red* and title will be displayed at the top of the scatter chart.
- *plt.show ( )* is a function which is going to show the scatter plot created by the code snippet.

The data points for the two separate quantities are represented by dots in a scatter plot, also known as a scatter chart or scatter plot. The values for each individual data point are given by the exact position of each dot on the vertical and horizontal axes. Connections between elements can be observed using scatter plots. Each small circle represents one tree, and the height (calculated in metres) and width (measured in centimetres) of each point correlate to the tree. The graph indicates that there is normally a strong beneficial relationship between a tree's height and diameter. Also, an outlier points that contrasts drastically in diameter from the other trees may be noticed. This tree is rather small given its overall size,

which might need additional investigation. Using scatter plots, correlating correlations are often identified. In such cases, we want to know what an appropriate projection for the high value would be, given a specific horizontal value. It is normal to see the independent variable portrayed through the axis that runs horizontally and the variable that is dependent represented by the vertical axis. A number of plenty of ways for identifying interactions between variables, namely either favorable or adverse, either good or bad, and straight or nonlinear. Further trends in information can also be seen with a scatter plot of the data. Based on how tightly sets of points crowd together, one can group data points into groups. likewise scatter plots can reveal any outliers or odd gaps in the information being examined. In the task of creating user personas, for instance that these can be handy if we want to split the data into multiple categories (YI, 2024).

6. Click on the *Run Script* and then the scatter chart is displayed (**Figure 4.20)**

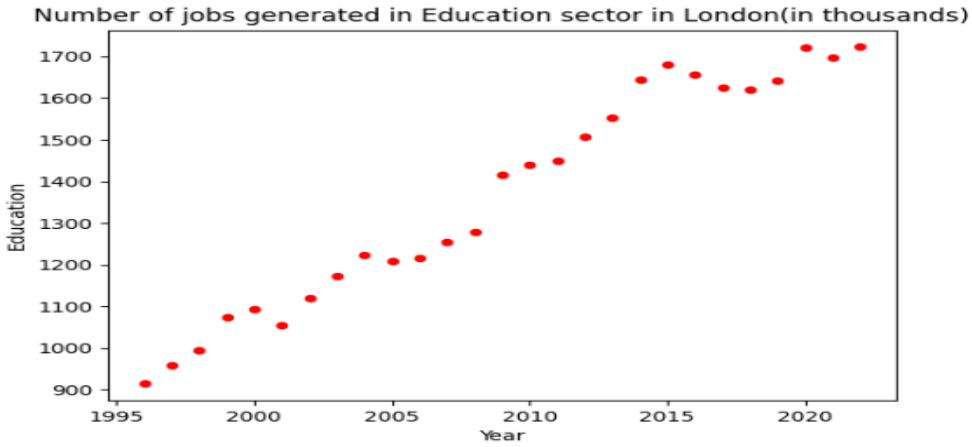


Figure 4.20 Scatter chart showing number of jobs in Education sector generated in London

7. In the same manner *line chart* is created with the code snippet.

```
Python script editor
⚠ Duplicate rows will be removed from the data.
1 import matplotlib.pyplot as plt
2 dataset.plot(kind='line', x='Year', y='Construction', color='blue',
3                title='Jobs generated in Construction sector in London(in thousands)')
4 plt.show()
```

Figure 4.21 Code to create the line Chart

8. Run the script and the *Line chart* is created **(Figure 4.22)**.

Figure 4.22 Line chart for the Construction sector in London

4.5 Building a Machine learning model – ARIMA in Python using Power BI

A*RIMA* stands for *AutoRegressive Integrated Moving Average*. It is extensively used for the prediction using *Time series.*
This model gathers insights from finding the patterns using time series. It is basically made up of three components which are:-

- *AutoRegressive*: It traces the correlation between the present and the lagged data. The lagged data is always represented by the letter 'p'.
- *Integrated:* It is applied to make the time series fixed. This is denoted by letter 'd'.
- *Moving Average:* It finds the relationship between different variables. It is referred by the letter 'q' .

The values of *p, d* and *q* are replaced by some integer numbers.
An inherent ARIMA the mechanism produced the time series. According on the processing observations, the values for p, d, and q must be effectively submitted. After applying the ARIMA model, the time series information needs to be reduced steady via difference. If the equation fits appropriately, its residuals need to be periodically distributed and unrelated. With the intent of simulating time series data for predictions along with other applications, the ARIMA model gives a structure and customisable method. (Awan, 2023).

ARIMA modelling underpins the main business planning, purchase, and production processes in most industry organizations. Projection erroneous will have a negative effect on the whole supply chain, or any company's atmosphere for that matter. Thus, it's crucial to have precise projections to reduce cost and ensure accomplishment. The approach that is most common is to establish it. In other words, take the old value from the fresh one. Several differentiations could be required, dependant on the series' depth. A time series was intended to grow stationary with the process of divergence. Nevertheless, one requires caution to avoid minimizing the series. Due to the chance that an over-differenced trend is still stationary, this could have an influence on the parameters used for the model. The minimal amount of differentiation needed for creating a near-stationary dataset that moves around a defined average and the ACF diagram quickly approaches negative is known as the "title order of variance." The series require additional that vary if the autocorrelations are good for ten or more delays. On the flip side, the series is probable over-differenced if the lag 1 synchronization is very unfavourable. If the effects that result from the intervening lags are taken out, the relation among the series and its lag can be described as partial redundancy. In this way, PACF models illustrates the pure correlation that occurs between a sequence and a lag. It is easy to choose and true that lag must happen in the AR term (Prabhakaran, 2024). Consequently, the minimum amount of differentiation required for maintaining the series steady can be determined by the value of d. Additionally, d is 0 if the period series is currently steady.

Steps to build machine learning model ARIMA by using Python in Power BI are: -

1. Launch the Power BI desktop and open the existing report *python machine learning model* **(Figure 4.8)**
2. Open a new tab and named it as ARIMA-Machine Learning

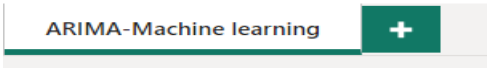


Figure 4.23 New tab created for ARIMA

3. Go to *Visualizations* and click on *Table.*
4. Select the column *Year* and *Construction.*
5. Click on *Get Data* and the select *Python Script.*

6. The code is to be written in *Python Script* as (**Figure 4.24, Figure 4.25, Figure 4.26)** and there is need to install the library which is *pmdarima* by using the command in command prompt *pip install pmdarima*.

Figure 4.24 Python code (1)

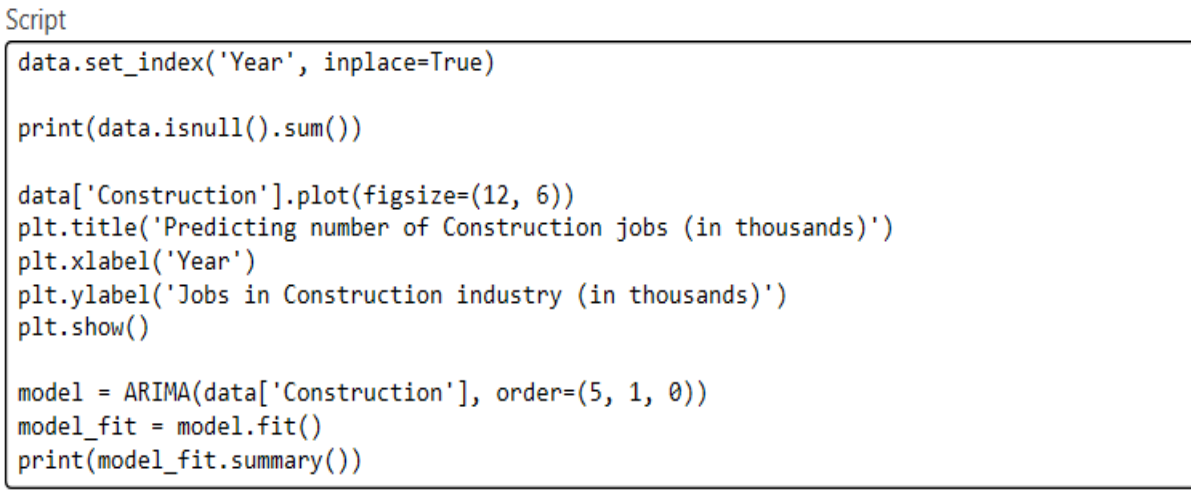


Figure 4.25 Python code (2)

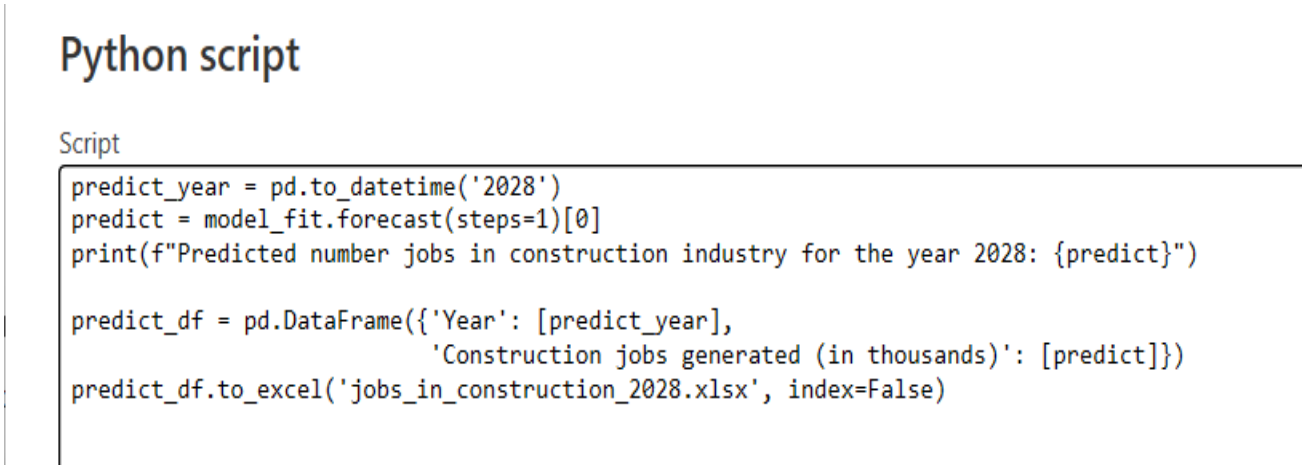


Figure 4.26 Python code (3)

7. Click on *OK* then the *Power BI* will pop up the window making the connection with the Python.
8. In the *Navigator window,* the two files are created *data* and *predict_df (***Figure 4.27)**

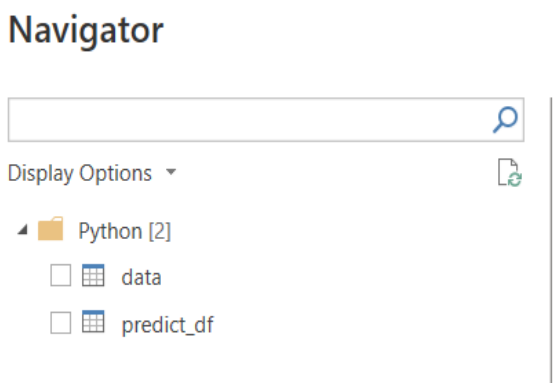


Figure 4.27 File created by Python Script

9. Click on predict_df and then the navigator window displays output *1254 thousand* jobs generated in *London* for the year *2028 (***Figure 4.28)**

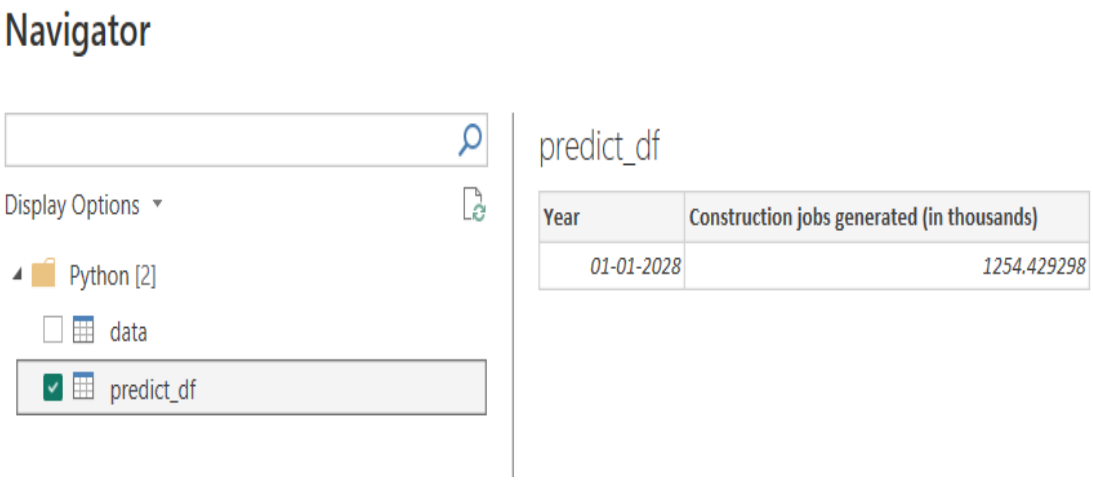


Figure 4.28 Predicted number of jobs for construction industry

10. Click on *Load*, the data will be displayed on the *Data Area (***Figure 4.29)**

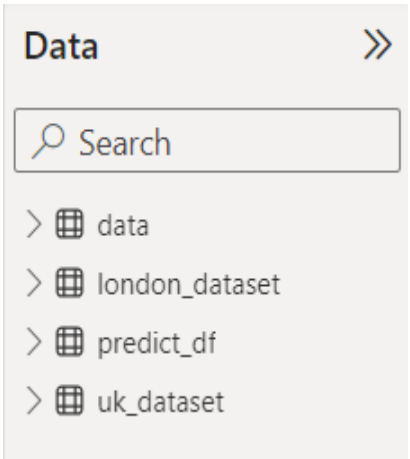


Figure 4.29 Predicted data popped up

11. The predicted value can be visualized in Power BI is represented by *Clustered column chart*

Figure 4.30 Predicted value visualized

12. It is also going to plot the line graph (**Figure 4.31)**

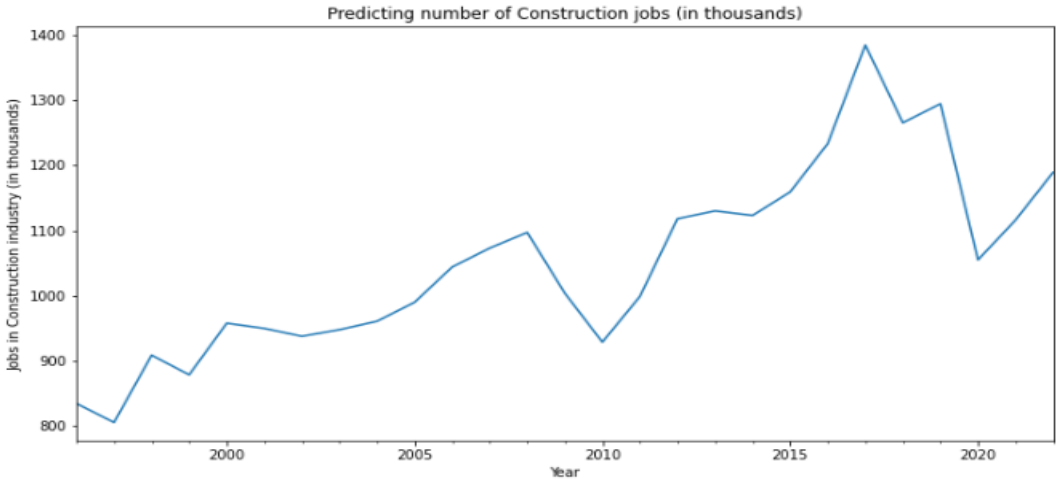


Figure 4.31 Line graph plotted by Python script

Code description

Code	Description
import pandas as pd	Importing the *pandas* library and creating its object *pd.*
import matplotlib.pyplot as plt	It is a library with which I created visualisations of the data in python and *plt* is its object.
from statsmodels.tsa.arima.model import ARIMA	This is importing ARIMA model for predicting the future data
file_path="C:/Users/mylap/OneDrive/Desktop/jobs 05dec2023.xlsx"	A variable called *file_path* is created which is storing my dataset file jobs05dec2023 which is excel format
sheet_name = 'london_jobs_data'	Since my dataset has 2 sheets *london_jobs_data* and *uk_jobs_data*, therefore a variable is created *sheet_name* which is going to hold the sheet *london_jobs_data*
data = pd.read_excel(file_path,	A variable *data* is created which is reading the excel file with *pd.read_excel*

sheet_name='london_jobs_data')	function which takes 2 arguments: the location of the file and sheet_name which is *london_jobs_data*
print(data.head())	It is going to print the first 5 rows of the dataset along with the column name
data['Year'] = pd.to_datetime(data['Year'], format='%Y')	Firstly the *Year* column is accessed and then it is converted to date time format because otherwise the Interpreter would consider it as a string. *%Y* represent the 4 digit number is year
data.set_index('Year' , inplace = True)	It specifies in the It specifies in the *Year* column if the modifications are made then the old values must be replaced by the new values.
print (data.isnull( ).sum( ))	With this code, it is going to calculate the sum of empty values in each attribute or column of the dataset
data['Construction'].plot(figsize=(12, 6))	This will create the plot for the Construction data column with 12 inches as width and 6 inches as height.
plt.title('Predicting number of Construction jobs (in thousands)')	With the *plt* object of *pyplot,* I gave the title to the chart that is going to be plot in the argument.
plt.xlabel('Year')	With the *plt* object, I set the label of *x axis* as the *Year* column.
plt.ylabel('Jobs in Construction industry (in thousands)')	With the *plt* object, I set the label of y *axis* as the value defined in the column.
plt.show()	By using the *plt* object, the chart will be plotted
model = ARIMA(data['Construction'], order=(5, 1, 0))	A variable called *model* is created which is using the *ARIMA* function that takes *Construction* column as its first argument and *order* has 3 values *5,1,0.* Therse three values in the ARIMA model function represents the AutoRegressive, Integrated and Moving Average respectively.
model_fit = model.fit()	I created a *model_fit* variable which storing the values of predicted values created by ARIMA model
print(model_fit.summary())	This line of code is going to print the information of technical nature by ARIMA model
predict_year = pd.to_datetime('2028')	A variable name *predict_year* has been generated which is converting the year value to date and time format and storin *2028* as an argument which is the year that I want to predict.
predict = model_fit.forecast(steps=1)[0]	A variable *predict* is storing the values of the fitted forecast values
print(f"Predicted number jobs in construction industry for the year 2028: {predict}")	The argument *f* in the *print* function is a string of formatted nature and it is going to display the message by taking *predict* as an another argument.
predict_df = pd.DataFrame({'Year': [predict_year], 'Construction jobs generated (in thousands)': [predict]})	A variable *predict_df* is created which is storing the dataframe by taking arguments as the value to be printed and *predict* which contains collection of the predicted values .
predict_df.to_excel('jobs_in_construction_2028.xlsx', index=False)	In this code, the output is finally created and exported to excel format with the name mentioned in the argument.

CHAPTER -5 TREND ANALYSIS

Trend-1

The *Administrative & support service activities* sector in *London* shows a massive decline in the year *2020* **[Figure 5.1]**. The number of jobs generated in generated in year *2019* were *2448 thousand* and in *2020* were *2178 thousand*. This means it leaves *270 thousand* people unemployed and thereby the employment rate is reduced by *11.02 %*. It was during lockdown in pandemic that *24 %* of the *United Kingdom* workforce jobs were at stake. The *UK* announced lockdown on *23 March 2020* due to the massive affect of *Covid-19* corona virus worldwide. There were plenty of business that were shut down as there were significant loss of clients. As per *McKinsey's* midpoint scenario, the *UK GDP* was reduced by *9 %*. The pandemic put 7.6 million jobs at risk which included permanent layoffs and reduction in the working hours. The one who were working out of which *50 %* are working at the rate of less than *£10* per hour (Terra Allas, 2020).

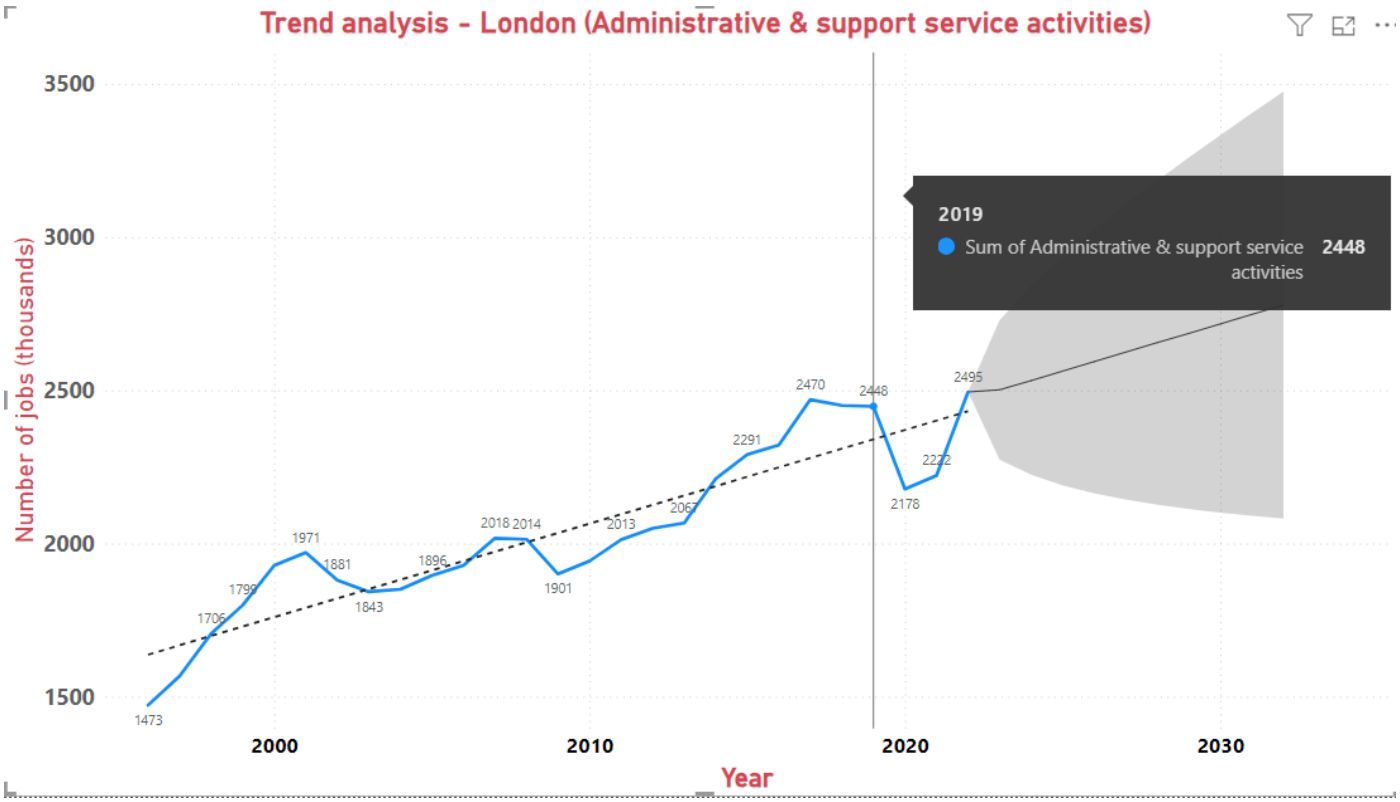


Figure 5.1 Line chart showing downtrend between the year 2019 and 2020

Trend – 2

In London, the *Administrative & support service activities*sector *displays* a significant high [**Figure 5.2]** as the number of jobs generated in year *2022* were *2495thousand* as compared to *2222 thousand* in year *2021*. This means *273 thousand* more jobs generated in year *2022* which is high by *12.28 %*. The primary reason for this hike is the lifting of restrictions after Covid-19 that created a boom in the job market. This displayed the sign of recovery from unemployment created by the Covid-19 restrictions. Eventually, with this sudden increase in number of jobs put the life on the track as the employers started their hiring process. Moreover, in some scenarios there were more jobs than the candidates which gives the candidates to negotiate on their salary packages and other factors. In London, only in month of May, 51700 jobs were published. Apart from that, Virtual work or work from home concept started which makes flexible for the both the employers and employees to work from their home location. This

led to saving of transportation time and cost. The virtual job sector created more jobs which led to the boom in job sector as well (Susan Lund, 2021).

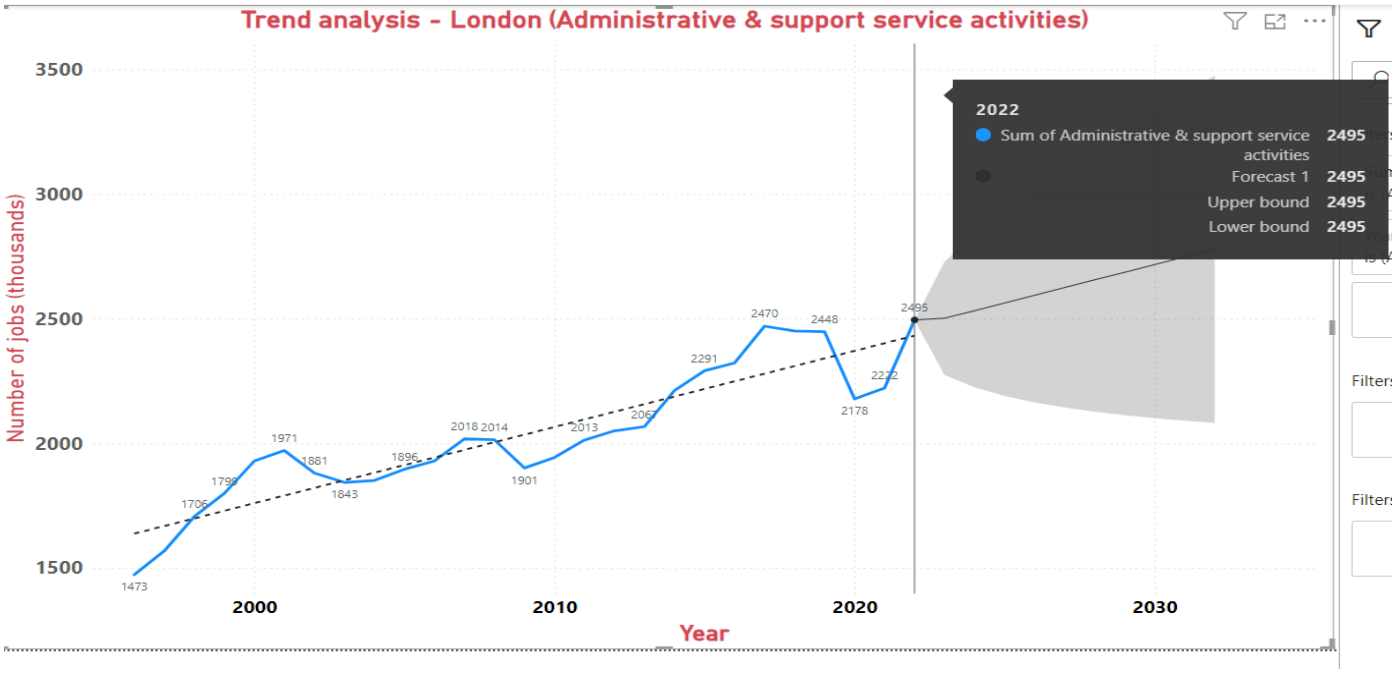


Figure 5.2 Line chart displays uptrend from year 2021 to 2022

Trend - 3

The *Construction* sector in *London*had faced a decline [**Figure 5.3]** in the year *2020*. Jobs generated in the *2020* were *1055thousand* which were *239 thousand* less than the year *2019* which produced *1294 thousand* jobs in the construction sector. The primary reason is the lockdown due to the pandemic. During that time no new orders and projects were started as the cost of construction raw material, fuel, energy was sky rocketing. This eventually increase the cost of construction due to which the developers and builders postpone their plan till the situation becomes stable and lifting of the ban from the lockdown. According to the survey conducted by *Construction Online*, almost *38 %* of the construction companies faces financial difficulty (Lowe, 2023).

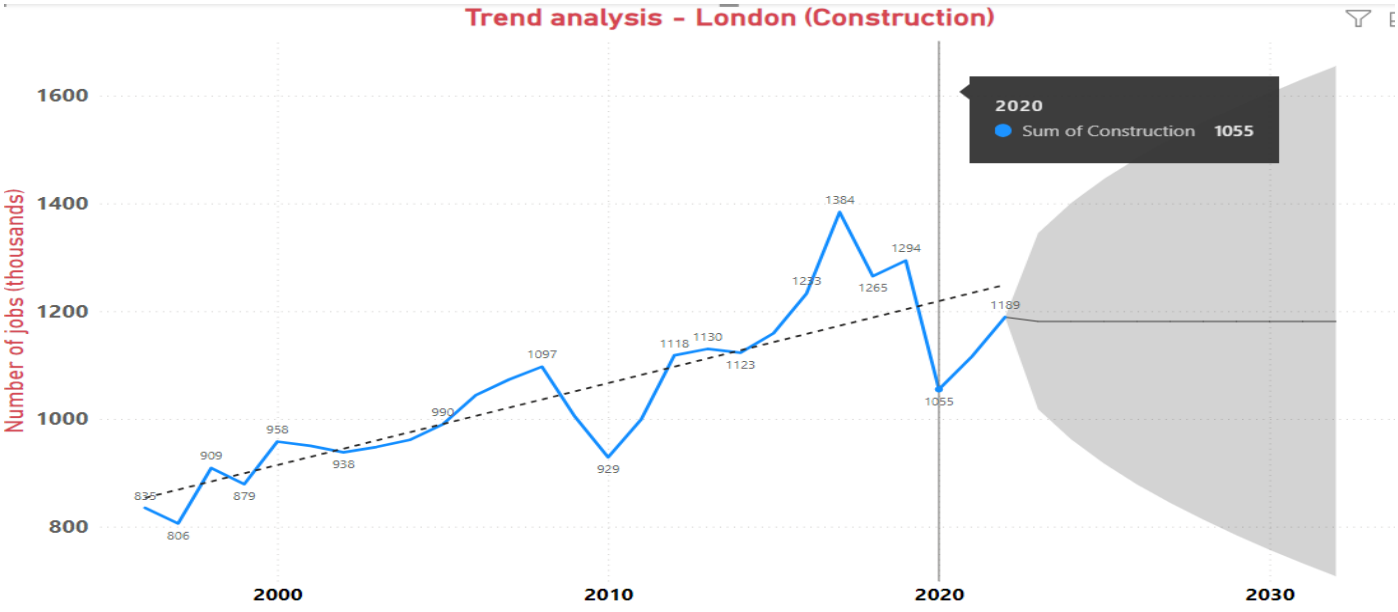


Figure 5.3 Line chart displays decline in year 2020

Trend - 4

In the year *2022*, the employment was at a massive extent in *Construction* sector as the number of jobs generated were *1189 thousand* which were *134 thousand* [**Figure 5.4]** greater than the year *2020*. This shows the uptrend by approximately *12.70%.* Even though, there was another variant of *Covid-19, Omicron* came in into effect, but it didn't had a great impact on the economic activity as the score of the people are getting vaccinated. The highest number of construction activity happened in the month of *February 2021* which was *1.6%* growth that is approximately *£14.6 billion*. Construction activity in London was up by *18 %* which is quite higher as compared to overall *UK* (Lowe, 2023).

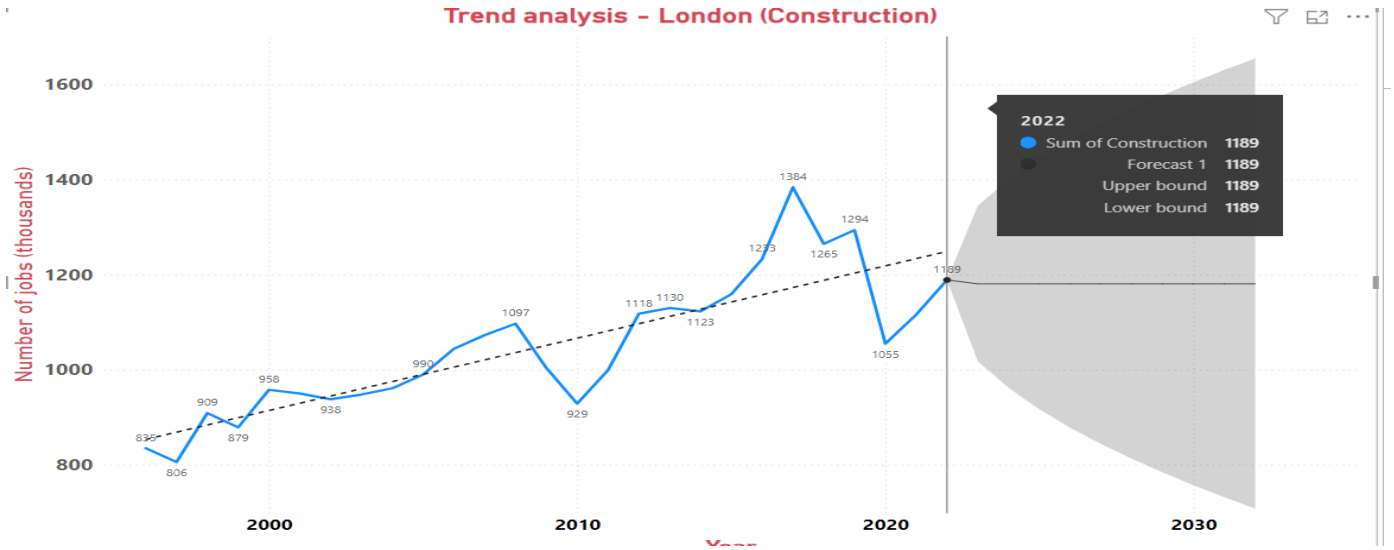


Figure 5.4 Line chart displays uptrend in year 2021

Trend - 5

The *Information & Communication* sector experience an uptrend in the year *2022*[**Figure 5.5]** that created *2196 thousand* jobs which were *12.26 %* higher than the year *2021* that only generated *1956 thousand* jobs. As per the data for *Office for National Statistics*, more than *1.9 million* jobs filled in UK which were *325 thousand* more than *2019*. As compared to *2022*, this number was increased by *4.1%.* The *Digital sector* filled up *5.6 %* of the total jobs. The Computer*programming* and *consultancy* shows *33 %* growth (National statistics, 2023).

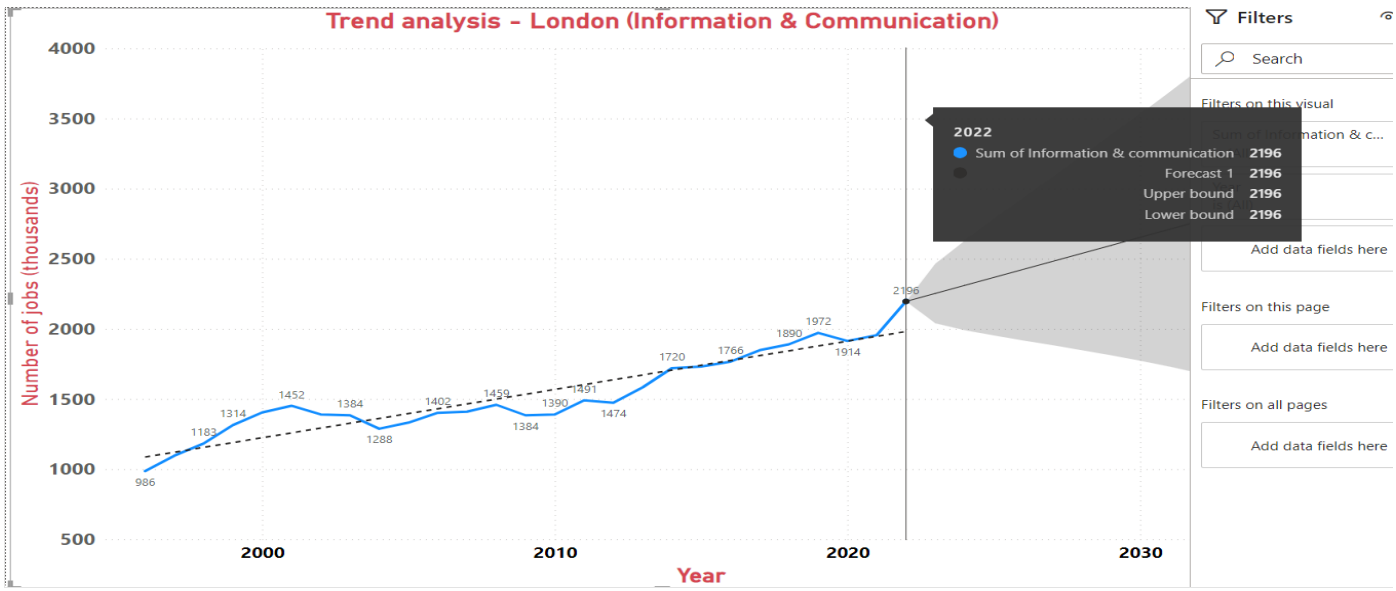


Figure 5.5 Line chart shows uptrend in 2022 in Information & Communication sector.

Trend – 6

Human health & social work activities sector produced vast amount of employment opportunities with a massive difference. As per the dataset, the number of jobs generated in *2019* in this sector were *2369 thousand* as compared to *2541 thousand* in *2021 [Figure 5.6]* that is basically more than *173 thousand* more with respect to year *2019* which constitutes around *7.26%* increase.

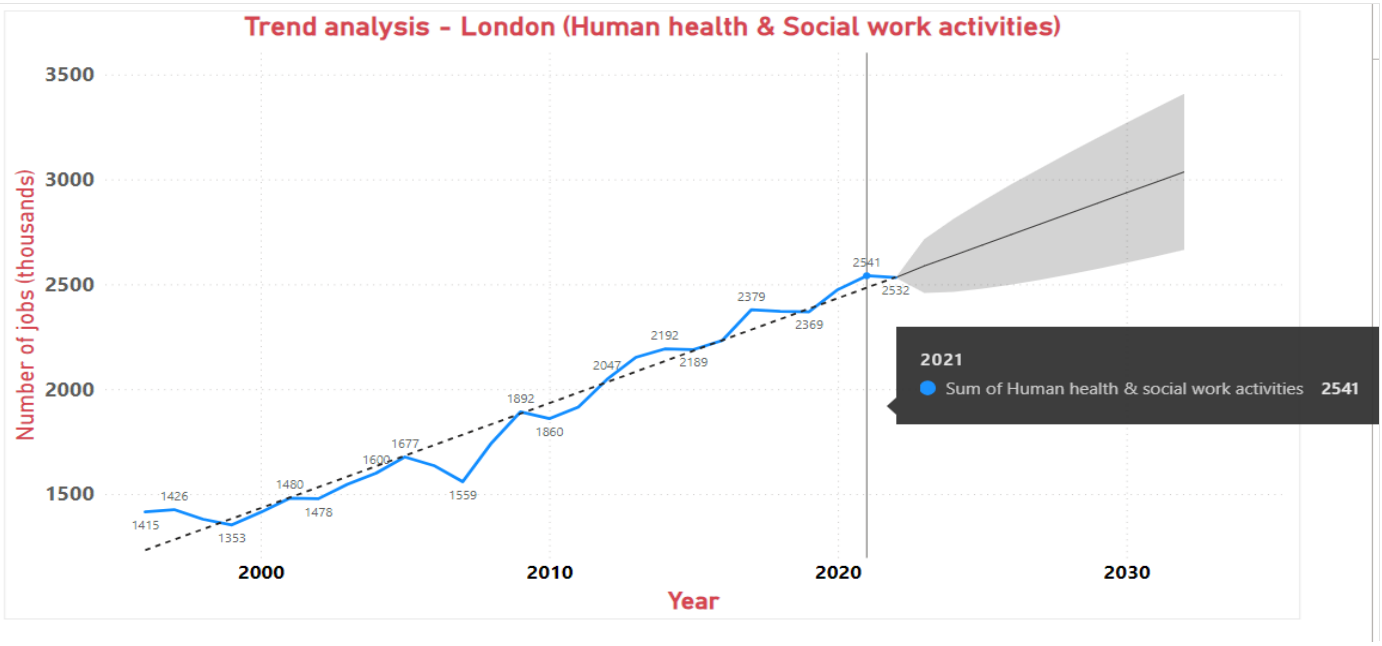


Figure 5.6 Line chart shows the uptrend in year 2021

Trend - 7

*Wholesale & retail trade; repair of motor vehicles and motorcycles*sector in UKgenerated almost *4.9 %* less jobs [**Figure 5.7]** as compared to year *2019*. In *2019*, it produced *20018 thousand* jobs whereas in *2022*, this sector showcases the downtrend by producing almost *19025 thousand* jobs which is less with quite a big difference of *993 thousand* jobs less jobs. There have been *63000*amount of jobs that were lost in the *2022*. The Covid-19 had a massive bad impact on this sector. In *Quarter 2* of *2022*, there were *14 %* less jobs as compared to *Quarter 1* of *2022*. There were 510 retail insolvents in *Quarter 2* of *2022*which was 31 % higher than *Quarter 1* of *2021*. Apart from that, the retail labour activity has been reduced by 0.4 % in *Quarter 2* of *2022* (KRIS HAMER, 2022)

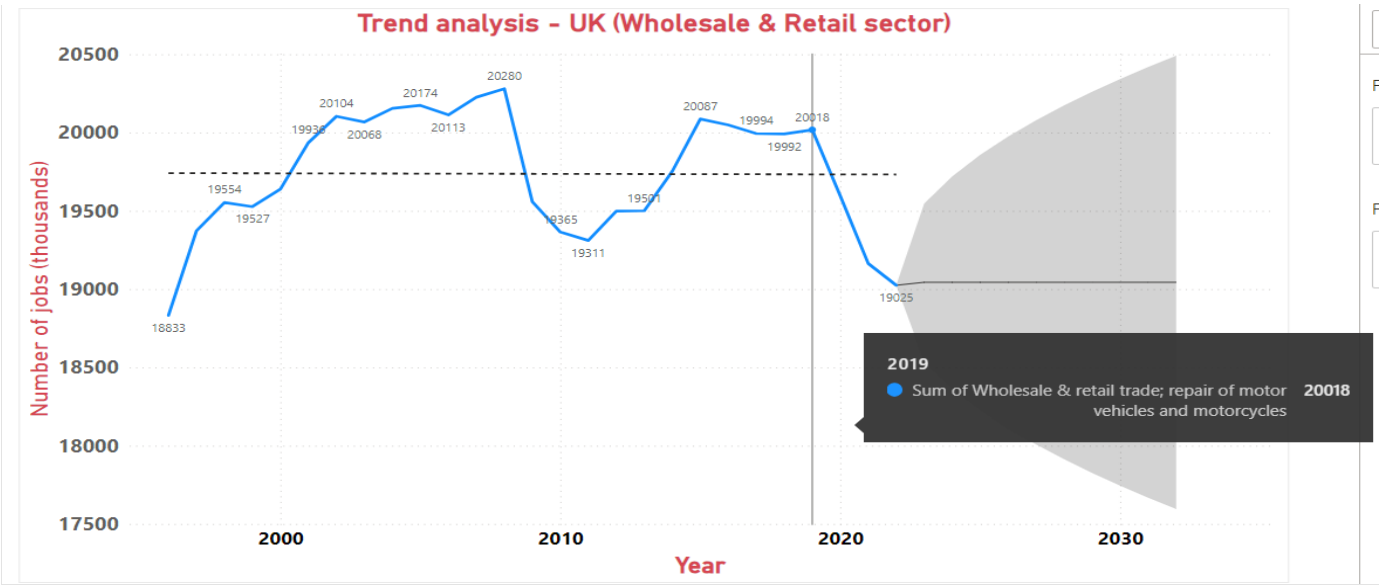


Figure 5.7 Line chart shows downtrend in Wholesale & retail sector from 2019 till 2022

Trend – 8

Human health & Social work activities sector have seen a massive uptrend in the year *1999* when the number of jobs generated were *16503 thousand* **(Figure 5.8)** which were *41 %* higher than in the year *1998* that constitutes around *11694 thousand* jobs. This is basically the rise of *4809 thousand* jobs just in a span of one year. As per the article when the Labour government came into power, there were lack of Nursing staff across the UK. The government needed more people in health sector due to the *Influenza* crisis in the last winter. The international staff was hired at a big rate to meet the demand for health sector. Another reason to hire more people was there are plenty of nurses are on the age line of *50*. *NHS Executive* created recruitment and retention department to hire and retain people at a big scale (James Buchnan, 2000).

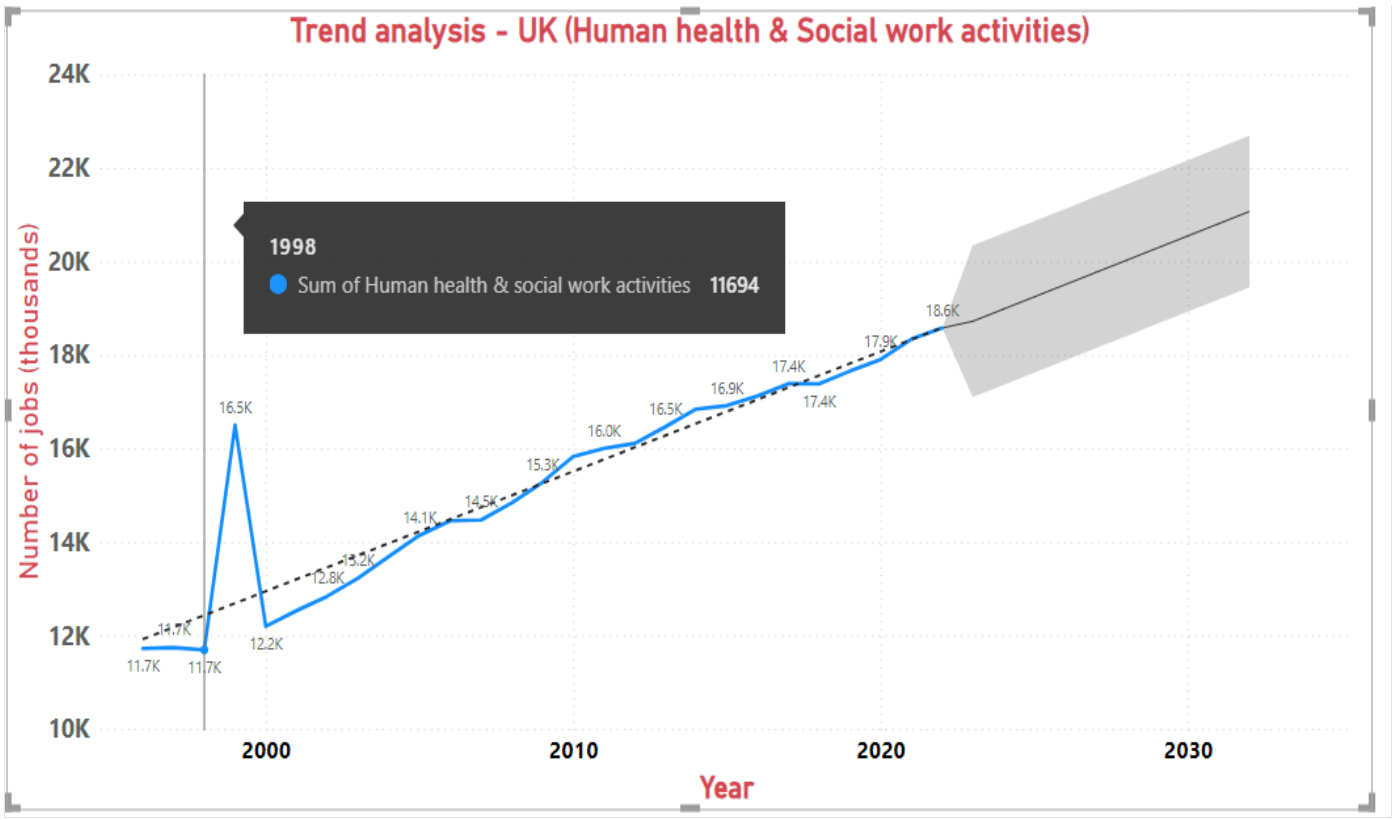


Figure 5.8 Line chart shows uptrend from year 1998 to 1999 in UK for Human health & Social activities

Trend – 9

The *Administrative & Support service activities* sector had gone through sudden downfall by making too many people unemployed as the recession started in *2008* which created *2014thousand* jobs **[Figure 5.9]** that dropped to *1901thousand* jobs which is approximately *113thousand* less jobs created that accounts to approximately *-5.6 %*. By the end of *Quarter2*, the *GDP* fell by *6 %*which is lot worse than the recession of *1980's* and *1990's*. Most of the hiring was almost stopped for the next 18 months.

This eventually led to cutting of overtime work and normal work hours. The full time working professionals are converted to part time, laid off or temporarily laid off from the work. The normal working hours dropped by *4 %*. As per the *Labour Force Survey (LFS)* , the unemployment rate for 16 + years was increased by *2.6 %* as compare to the *1980's* or *1990's* (Paul Gregg, 2010) .

By the starting of *2008*, the rate of economic activities was next to nil. People lost jobs and at the end of *2011*, there were 2.7 *million* people were looking for work. The quarter wise unemployment rate was reached at *8.4 %* which was highest since *1995*. (Velma Loh, 2018)

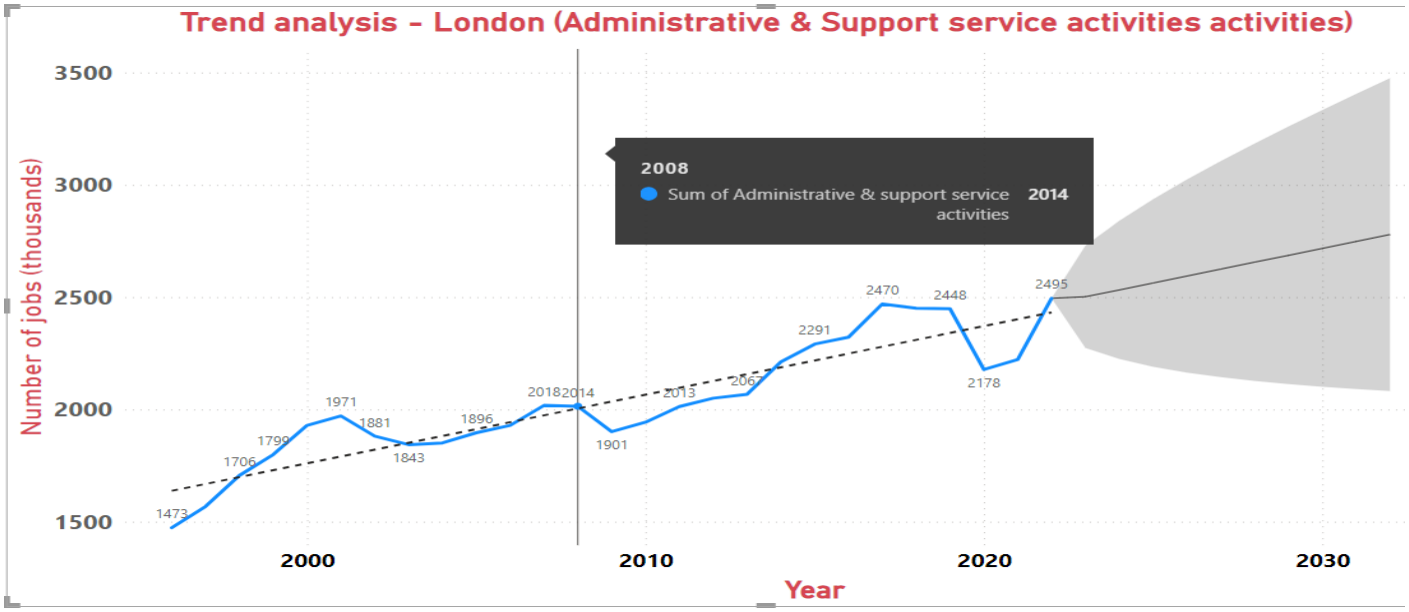


Figure 5.9 Line chart shows downtrend from year 2008 to 2009 in London (Administrative & support service activities)

Trend – 10

The *Construction*sector in *UK* experiences the sudden downfall starting from year *2008* and goes till *2010*. The number of jobs generated in the year *2008* was *9312 thousand* **[Figure 5.10]** and then the Recession came into existence and continued till 2010 that only generated 8312 thousand jobs. So, it is highly probable that the building market is not already firmly in recession, according to Business Insight economist Dean Archer. The construction industry appears to be facing a protracted and extremely challenging period. This underscores our view that a downturn of the entire economy in the second part of 2008 is likely than not.

This is one of the biggest unemployment rates as it constitutes 1 million jobs loss which is eventually 10.73 % less. The construction was as low as 18.7 %. The construction activity was at its lowest form during this period of time. Charted Institute of Purchasing Supply, Construction PMI drops to 36.7 in the July quarter of 2008. As per the report anything below the rate of 50 indicates, the sector is shrinking (The Guardian, 2008) .

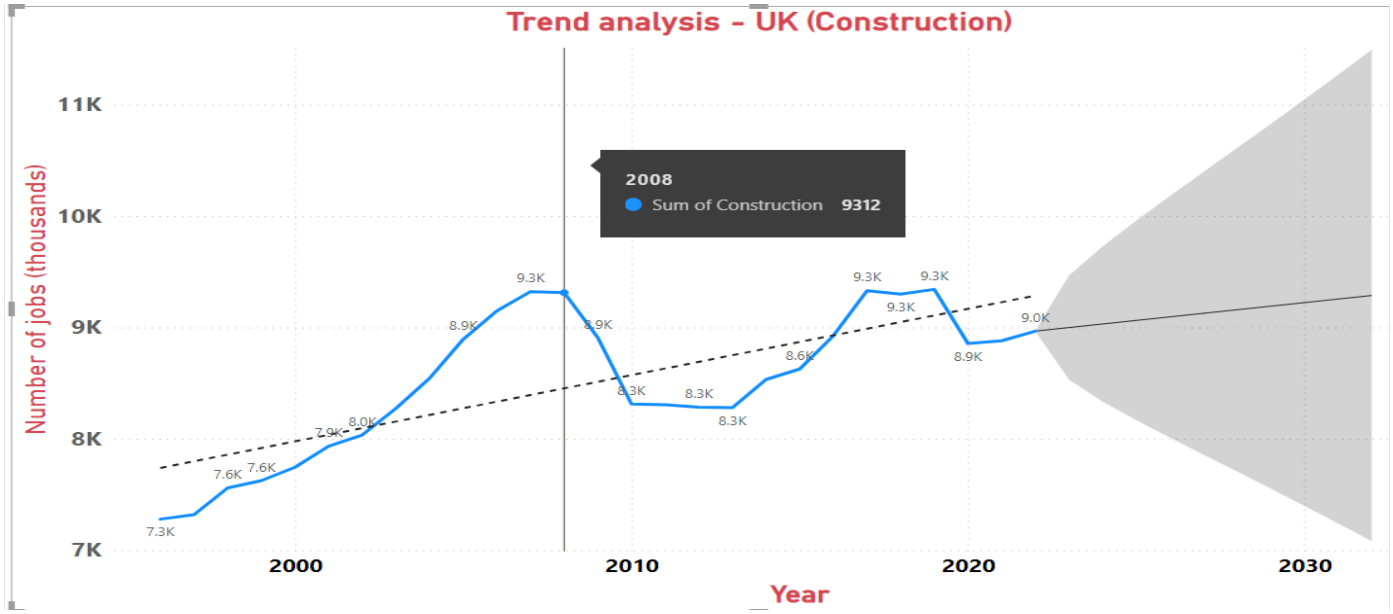


Figure 5.10 Line chart showing downtrend in UK – Construction sector

Trend – 11

The *Education* sector in *London* shows slump as far as the job generation is concerned. The number of jobs generated in *2015* were *1680 thousand* as compared to *1621 thousand* jobs in *2018* **[Figure 5.11].** This is a sudden fall by *3.5%* which is approximately *59 thousand*jobs less generated. Overall, this sector is in continues uptrend from the year *1996* till *2022*.

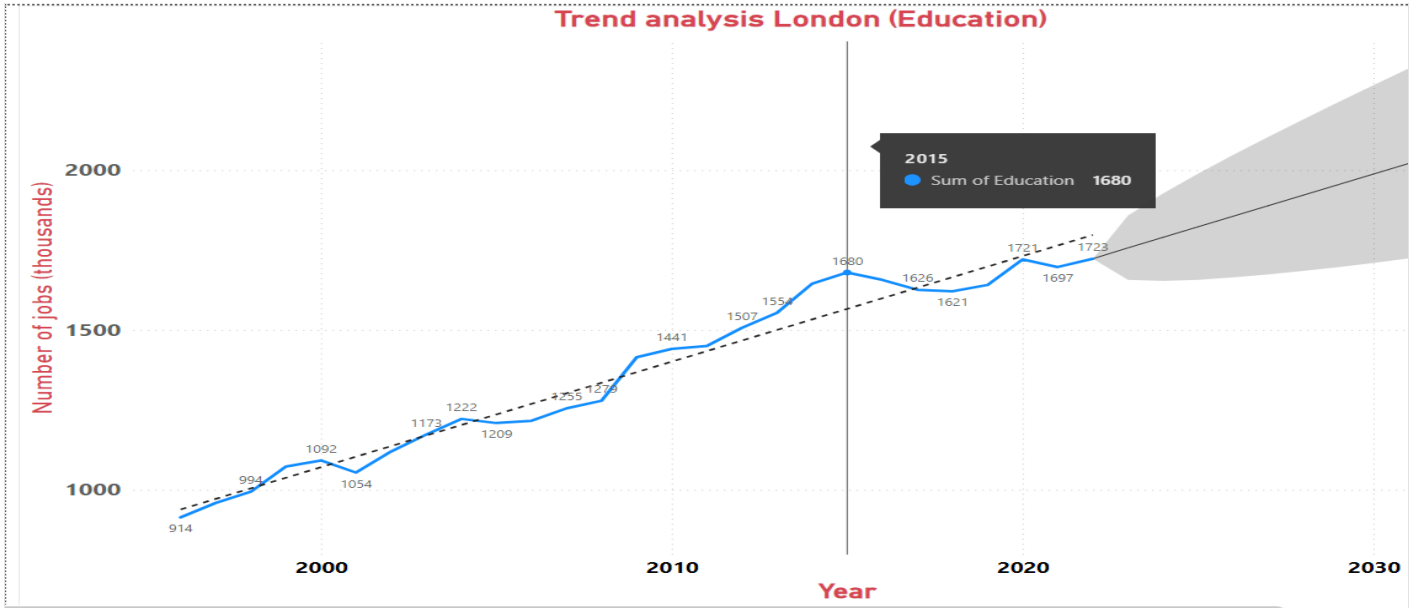


Figure 5.11 Line chart shows downtrend in Education sector from 2015

Trend-12

The *Information & Communication* sector in *London* shows enormous improvement from 1996 to 2001**[Figure 5.12]** as far as the jobs creation is concerned. In *1996*, the number of jobs yield in this sector was*686 thousand* which is approximately *47.26* mammoth growth.

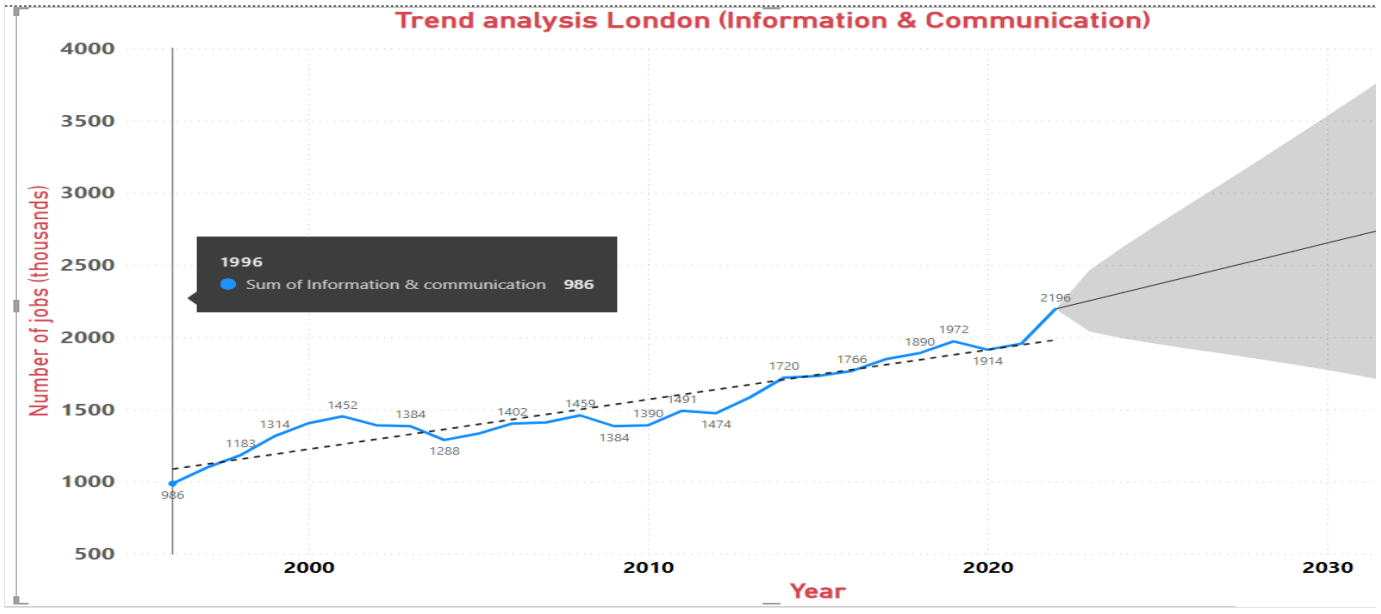


Figure 5.12 Line chart shows downtrend in Information & Communication sector from 1996

CHAPTER-6 COMPARATIVE STUDY

Study-1

The London's job data generate various insights by having a look at the Pie chart **(Figure 6.1).** During the computation of *Average* for the dataset it has been found that the*Wholesale & retail trade; repair of motor vehicles and motorcycles*has generated most number of jobs with the average number of jobs nearly around *24.48 %* of all the jobs created. However, the *Construction* sector generated the least number of jobs with just *10.09 %.*

The second most job generated sector average wise is *Administrative & support service activities* with the average of *19.52 %.* The *Human health & social work activities* sector generated jobs with the average of *18.07%* from *1996* to *2022.* The *Information& communication* sector generated jobs with the average of *14.72 %.* The *Education* sector constitutes *13.2 %.*

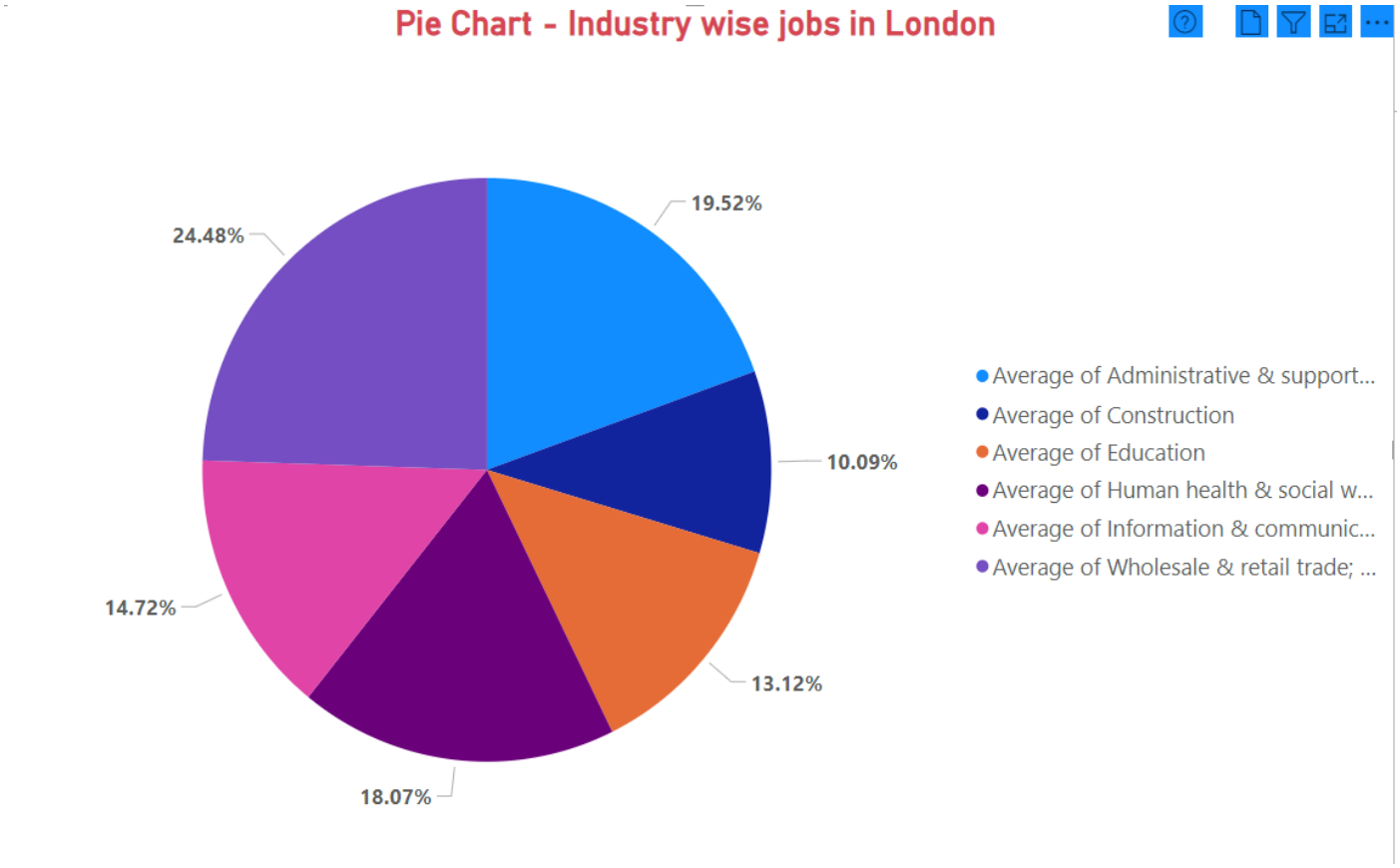


Figure 6.1 Pie chart shows industry wise average creation of Jobs in London

By looking at the average number of jobs created by sector in all of the UK (**Figure 6.2)**, it can be clearly seen that the*Wholesale & retail trade; repair of motor vehicles and motorcycles*has generated most number of jobs with the average number of jobs *28.55 %.* The least number of jobs has been created on an average is in *Information & communication* sector which is around *7.3 %.*

The *Human health & social work activities* sector accommodates the second spot having *22.07 %* share.*Education* sector achieved the average of *15.15 %.*It is followed by *Administrative & support service activities* sector with *14.61 %.* The second last placed is acquired by *Construction* sector which is around *12.31 %.*

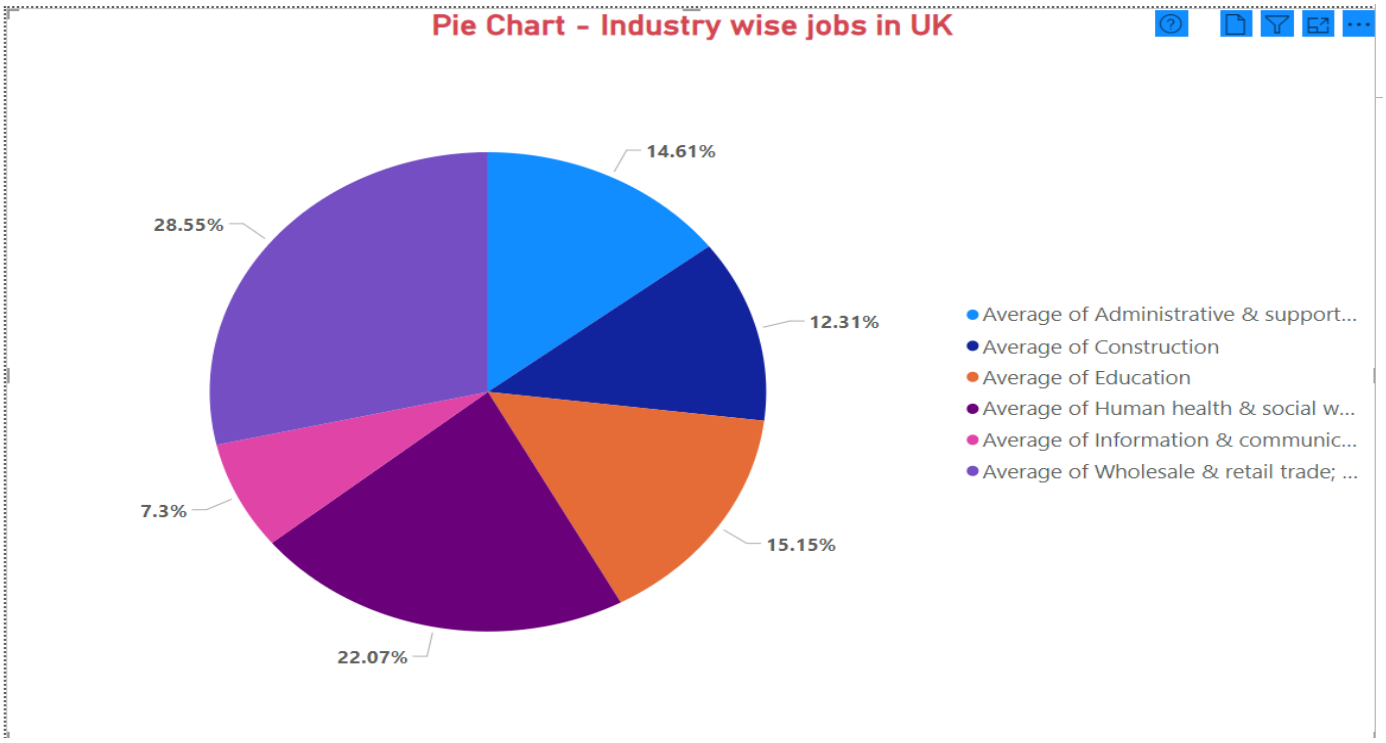


Figure 6.2 Pie chart shows industry wise average creation of Jobs in UK

Study-2

It is clearly visible in another study that all the four sectors of *Administrative & support services, Construction, Information & communication* and *Sum of Wholesale & retail Trade* experience sudden fall from year 2008 to 2009.

The consequences of the global financial crisis and following restrictions on spending have had significant consequences on job markets. Recession of 2008 led to unemployment at large scale. The works allocated were converted from full time to part time.

Between 2008 and 2014, job in the public sector reduced (by about 150,000), Most of the increase in zero-hour employment occurred in the private sector, however it's possible that this is also a result of the public sector contracting out operations. However, an equal number of staff in both the public and private sectors claim that they are underemployed (9% vs. 10%) or that the reason why they're employed freelance is because they were unable to secure full-time employment. Although average wages have not kept up with growth since 2008, the national minimum wage, at £6.50, provides a floor for hourly salaries. Also, a greater percentage of people receive compensation at the lowest level (Stephen McKay, 20).

.

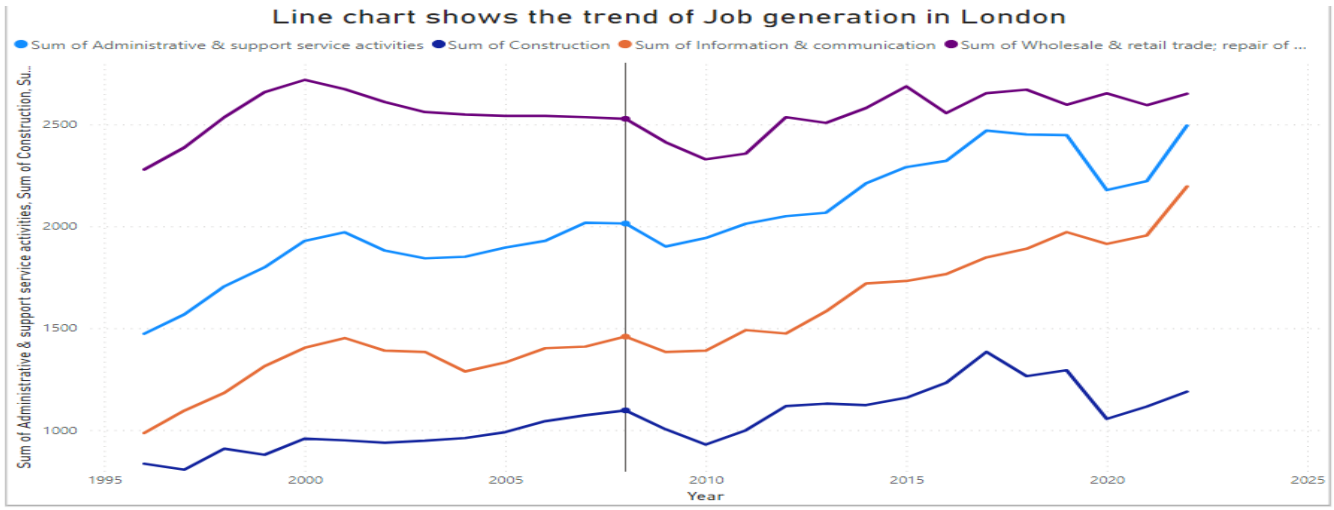


Figure 6.3 Line chart shows the similar downfall on the 4 sectors

Study-3

The line chart with *Time Series* forecasting conveys that there is continuous uptrend (**Figure 6.4)** in the job generation in London and UK.

Furthermore, it is showing uptrend in future as well.

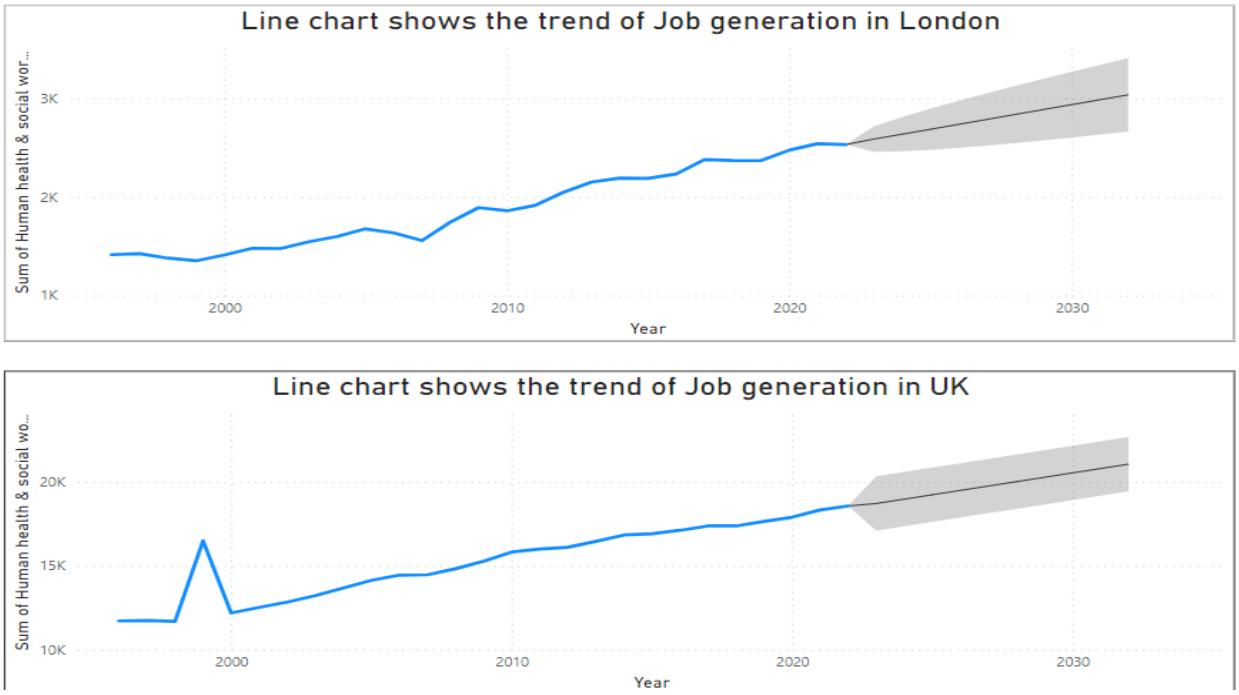


Figure 6.4 Line chart shows the overall uptrend in both London and UK

Study 4

The line chart in both the London and UK for education sector is also showing ongoing uptrend along with their future forecasting.

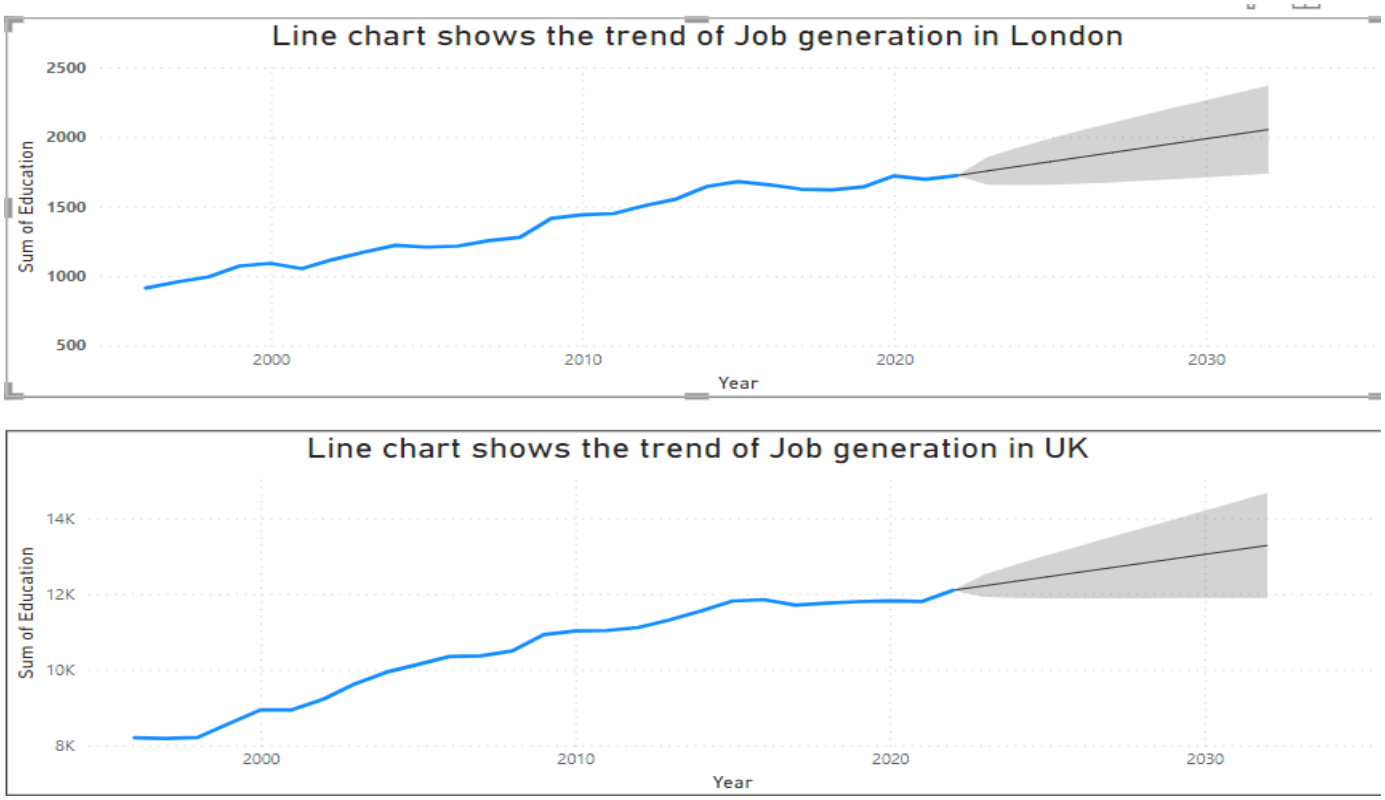


Figure 6.5 Uptrend in Education sector job generation

Study-5

The Line chart shows regular hike in the employment generation in the Information and communication sector in London and UK (**Figure 6.6)**.

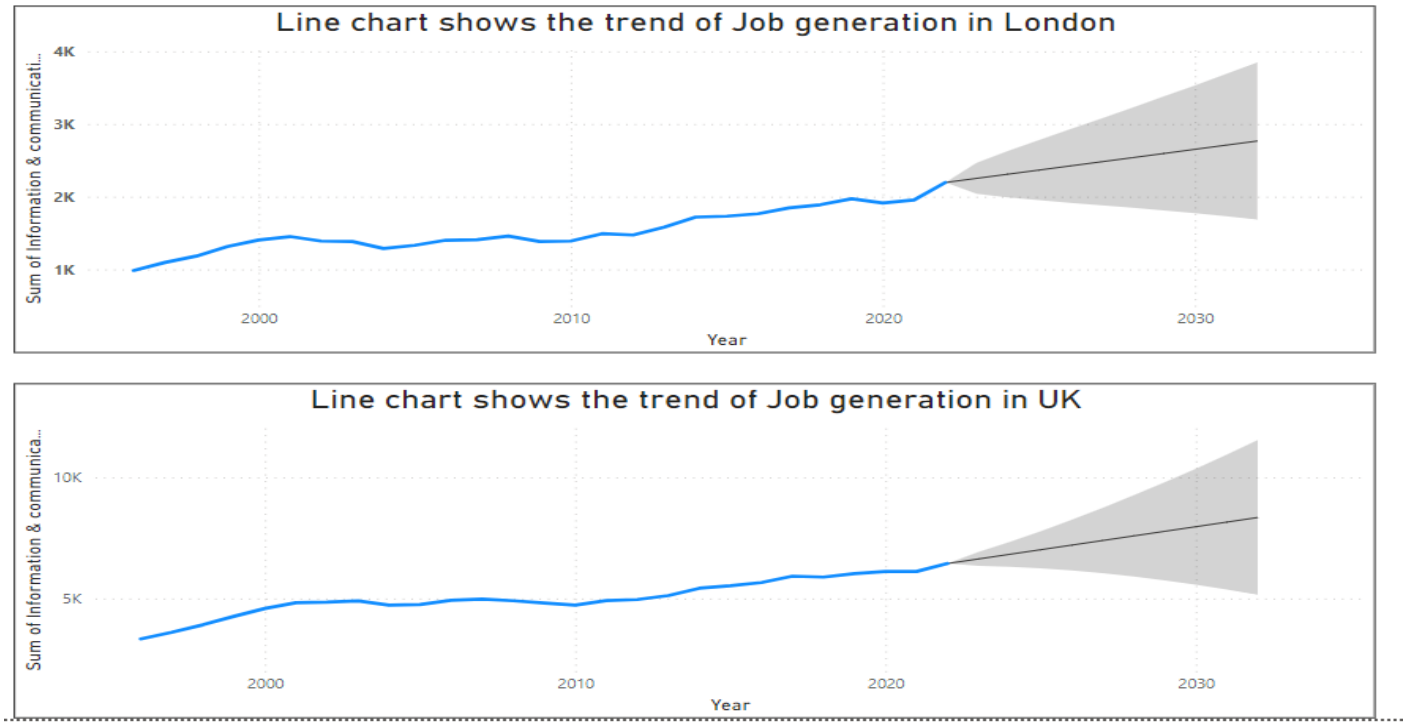


Figure 6.6 Hike in trend for Information and Communication sector in London and UK

Study-6

The Wholesale and retail sector shows the uptrend and then constant in future in London. However, in UK it is downtrend overall and then shows the slight downtrend as well.

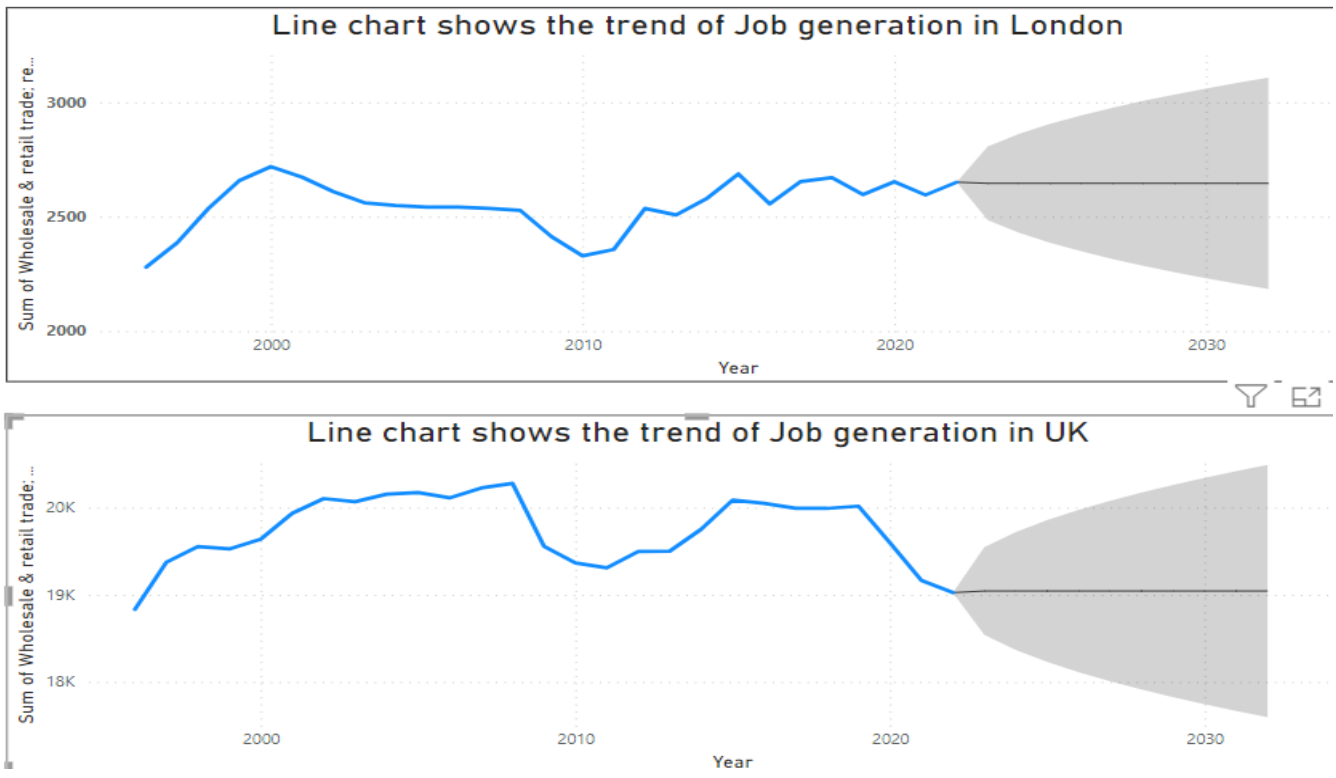


Figure 6.7 Trends in Wholesale sector for both London and UK

6.2 Limitations:

The limitation which would be more likely to occur is that the data being collected by government agency have been taken from plenty of surveys. Those surveys seldomly provide inconsistent data. The inconsistent data if been injected in the model results sometimes in improper forecasting. Secondly, the forecasting is based on old data which has been impacted by various factor for instance Pandemic, Recession, Brexit which no one know are likely to occur in future or not. While implementing the *Linear Regression*, which has only *Year* has an independent variable which remains increase constantly by 1 that eventually lead to question mark on the forecasting generated. The future will also be impacted by the use of Artificial intelligence and automation because many sectors are employing technology to do the task that human which may result in cutting of jobs or generation of new jobs.

6.3 Future Research Directions

In the future, I would like to add few more factors to make the insights better understanding and result conveying. The first thing I will do is to take the factors into consideration like GDP, inflation. This can be done by implementing new machine learning model techniques that will be developed over time. The other way around is to access the data with the help of real time systems. The advantage of real time system is that it keeps on updating at regular basic. This means, the analyst does not have to wait for too long for the latest data. Prediction and forecasting could be done with the blink of the eye. With the more advanced tools, I would like to understand bridge the gap between the education systems and the skill requirement.

CHAPTER 7: CONCLUSION

7.1 Summary of insights:

In my research, I tried my level best to address my research question by using machine learning models for the prediction. Moreover, I make sure the data remains concise and cleaned for the purpose getting accurate predictions. The Recession 2008 cause downtrend in the jobs in the sectors of Wholesale and trade, Administration, Construction and Information Technology. It is also found the health & social sector, Education and Information & communication are on regular high trend since from 1996 in both the London and UK. The Wholesale & retail sector in London and UK both face opposite trends as the UK is going slightly down whereas the London sector keep on growing. The insights generated from Linear regression and Time series models. The data from time series is taken into consideration because I am getting the prediction from the Historical data and trend patterns whereas the Linear Regression models was highly concentrated on the independent variable Year which is constantly increasing by 1 every year. In the research, the values generated from forecasting figures are also compared to highlight the difference.

7.2 Practical Implications with contributions to knowledge

This research on its part tried to contribute at a large scale for forecasting and predicting. In old studies, not too many models used to predict the information in the coming years. This study contributes by incorporating *Linear Regression*, *Time series* and *Machine learning model* in *Python.* The data was properly segregated and cleaned before the putting it for analysis. This was done to understand the data in a simpler way and created the bigger picture of the overall scenario. Apart from *Power BI*, the *Python scripts* are also incorporated for the objective of creating more authentic information that will be useful for the policy makers, recruitment specialists and government to provide a keen role making policies and training modules.

Furthermore, the trends in both the UK and London have been studied and represented by *Power BI* and *Python* visuals. The sudden hike or fall of certain trend is also justified with the reason from the authentic sources. The major reason for the sudden

change in trends were *Covid 19*, *Recession* and *Brexit.* It is also found which sectors will generating most and least number of jobs in London and UK overall using forecasting analysis. The forecasting of London and UK data has been studied using comparative study analysis to find insights, gaps and further recommendations.

7.3 In Nutshell

In the Conclusion, I would say London has more scope for jobs in every sector. To prove the same, it has been justified by the implementation of the modelling techniques in Power BI and Python. I gathered the data which will be beneficial for the concerned people to frame the policies and laws to support future job seekers in their quest for hunting the right job with right skill.

REFERENCES

Awan, A. A., 2023. *Time Series Analysis: ARIMA Models in Python.* [Online]
Available at: https://www.kdnuggets.com/2023/08/times-series-analysis-arima-models-python.html
[Accessed 13 April 2024].

BEERS, B., 2024. *Regression: Definition, Analysis, Calculation, and Example.* [Online]
Available at: https://www.investopedia.com/terms/r/regression.asp
[Accessed 11 March 2024].

GANTI, A., 2019. *What Is the Multiplier Effect? Formula and Example.* [Online]
Available at: https://www.investopedia.com/terms/m/multiplierеffect.asp#toc-the-keynesian-multiplier
[Accessed 21 February 2024].

geekforgeeks, 2022. *Life Cycle Phases of Data Analytics.* [Online]
Available at: https://www.geeksforgeeks.org/life-cycle-phases-of-data-analytics/
[Accessed 24 February 2024].

HAYES, A., 2022. *What Is a Time Series and How Is It Used to Analyze Data?.* [Online]
Available at: https://www.investopedia.com/terms/t/timeseries.asp
[Accessed 11 March 2024].

Higham, W., n.d. *2026: London's place in the.* [Online]
Available at: https://www.businessldn.co.uk/sites/default/files/documents/2018-05/London-2026.pdf
[Accessed 21 March 2024].

Homebase Team, 2023. *Workforce forecasting: what it is and how to predict future needs.* [Online]
Available at: https://joinhomebase.com/blog/workforce-forecasting/
[Accessed 24 February 2024].

IBM, 2024. *What is ETL (extract, transform, load)?.* [Online]
Available at: https://www.ibm.com/topics/etl
[Accessed 7 March 2024].

IMSL, 2021. *What Is a Regression Model?.* [Online]
Available at: https://www.imsl.com/blog/what-is-regression-model
[Accessed 13 March 2024].

James Buchnan, N. E., 2000. *Nursing numbers in Britain: the argument for workforce planning.* [Online]
Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1117947/
[Accessed 28 March 2024].

KRIS HAMER, H. D., 2022. *RETAIL JOBS REPORT,* s.l.: BRC.

Labour Market Team, 2023. *JOBS05: Workforce jobs by region and industry.* [Online]
Available at:
https://www.ons.gov.uk/employmentandlabourmarket/peopleinwork/employmentandemployeetypes/datasets/workforcejobsbyregionandindustryjobs05
[Accessed 26 February 2024].

Local Government Association, 2022. *Work Local: benefits of improving employment and skills outcomes.* [Online]
Available at: https://www.local.gov.uk/publications/work-local-benefits-improving-employment-and-skills-outcomes
[Accessed 4 March 2024].

Lowe, T., 2023. *Jobs at risk as construction activity falls by fastest rate since early months of the pandemic.* [Online]
Available at: https://www.building.co.uk/news/jobs-at-risk-as-construction-activity-falls-by-fastest-rate-since-early-months-of-the-pandemic/5121172.article
[Accessed 27 March 2024].

Lucina Rolewicz, D. B. P. C. L., 2024. *The NHS workforce in numbers.* [Online]
Available at: https://www.nuffieldtrust.org.uk/resource/the-nhs-workforce-in-numbers
[Accessed 25 February 2024].

Microsoft Build, 2024. *What is Power BI?.* [Online]
Available at: https://learn.microsoft.com/en-us/power-bi/fundamentals/power-bi-overview
[Accessed 12 April 2024].

National statistics, 2023. *Economic Estimates: Employment in DCMS sectors and Digital sector, January 2022 to December 2022..* [Online]
Available at: https://www.gov.uk/government/statistics/economic-estimates-employment-in-dcms-sectors-and-digital-sector-january-2022-to-december-2022
[Accessed 27 March 2024].

Neal Kilbane, N. M., 2017. *LONDON HOST BOROUGHS EMPLOYMENT FORECASTS,* London: Oxford Economics.

Otarb, M. S. m. T. v.-h. v.-k., 2022. *Run Python scripts in Power BI Desktop.* [Online]
Available at: https://learn.microsoft.com/en-us/power-bi/connect-data/desktop-python-scripts
[Accessed 3 April 2024].

Paul Gregg, J. W., 2010. *The UK Labour Market and the 2008 – 2009 Recession ,* London: Centre for Economic Performance, London School of Economics and Political Science.

Paul Gregg, J. W., 2010. *The UK Labour Market and the 2008 – 2009 Recession.* [Online]
Available at: https://cep.lse.ac.uk/pubs/download/occasional/op025.pdf
[Accessed 2 April 2024].

Peyton-Jones, T., n.d. *The Future of Work Jobs and Skills in 2030.* 1st Ed/03.14 ed. s.l.:UKCES.

Prabhakaran, S., 2024. *ARIMA Model – Complete Guide to Time Series Forecasting in Python.* [Online]
Available at: https://www.machinelearningplus.com/time-series/arima-model-time-series-forecasting-python/
[Accessed 23 March 2024].

Stephen McKay, R. S., 20. *The labour market before and after the recession.* [Online]
Available at: https://social-policy.org.uk/wordpress/wp-content/uploads/2015/04/McKay-Smith.pdf
[Accessed 24 March 2024].

Susan Lund, A. M. J. M. S. S. K. E. O. R., 2021. *The future of work after COVID-19.* [Online]
Available at: https://www.mckinsey.com/featured-insights/future-of-work/the-future-of-work-after-covid-19
[Accessed 28 March 2024].

Team Asana, 2024. *6 project constraints and how to manage them for project success.* [Online]
Available at: https://asana.com/resources/project-constraints
[Accessed 24 March 2024].

TechNikhil, 2024. *Introduction to Forecasting in Data Science.* [Online]
Available at: https://medium.com/@nkhb261/introduction-to-forecasting-in-data-science-0182f23f1f33
[Accessed 21 February 2024].

Terra Allas, M. C. D. V. H., 2020. *COVID-19 in the United Kingdom: Assessing jobs at risk and the impact on people and places.* [Online]
Available at: https://www.mckinsey.com/industries/public-sector/our-insights/covid-19-in-the-united-kingdom-assessing-jobs-at-risk-and-the-impact-on-people-and-places
[Accessed 26 March 2024].

The Guardian, 2008. *UK construction activity tumbles to record low.* [Online]
Available at: https://www.theguardian.com/business/2008/aug/04/construction.economics
[Accessed 28 March 2024].

tutorialspoint, 2022. *Python Tutorial.* [Online]
Available at: https://www.tutorialspoint.com/python/index.htm
[Accessed 28 March 2024].

Velma Loh, J. S., 2018. *The 2008 recession 10 years on.* [Online]
Available at:
https://www.ons.gov.uk/economy/grossdomesticproductgdp/articles/the2008recession10yearson/2018-04-30
[Accessed 29 March 2024].

Webber, A., 2024. *London and the South to see faster rate of jobs growth.* [Online]
Available at: https://www.personneltoday.com/hr/london-and-the-south-to-see-faster-rate-of-jobs-growth/
[Accessed 2 March 2024].

Will Higham, D. L., n.d. *2026: London's place in the UK economy.* s.l.:Social Market Foundation.

YI, M., 2024. *A complete guide to scatter plots.* [Online]
Available at: https://www.atlassian.com/data/charts/what-is-a-scatter-plot
[Accessed 17 March 2024].