## **ETL-Project: by Sarai Embaye, O.J. Ndebbio, & Terry Potter**

## **ETL Project Homework Assignment**

1. Project Description - The current economic market condition with respect to oil and gas prices is at an historical low. This has led to numerous lay-offs and furloughs within the oil and gas energy sector. I am curious to find out if the price of crude oil per barrel correlates to the unemployment rate on a national scale. See graph below which details 2020 YTD crude oil prices.



1. Project Goals - create a database tracking unemployment rates and crude oil prices on an annual basis.
2. Extraction (original data sources, formatting, source data description, and source description)
   1. Data Source 1, Labour: Labour market statistics, Jan 1955 - Mar 2020, CSV Format
      1. <https://data.oecd.org/unemp/unemployment-rate.htm>
      2. The Organisation for Economic Co-operation and Development (OECD) is an international organisation that works to build better policies for better lives. Our goal is to shape policies that foster prosperity, equality, opportunity and well-being for all.
      3. Unemployment rate is the number of unemployed people as a percentage of the labour force, where the latter consists of the unemployed plus those in paid or self-employment. Unemployed people are those who report that they are without work, that they are available for work and that they have taken active steps to find work in the last four weeks. When unemployment is high, some people become discouraged and stop looking for work; they are then excluded from the labour force. This implies that the unemployment rate may fall, or stop rising, even though there has been no underlying improvement in the labour market.
   2. Data Sources 2, Spot Crude Oil Price: West Texas Intermediate (WTI), Dollars per Barrel Not Seasonally Adjusted, Jan 1946 to Mar 2020, CSV Format (Note: Spot price is current market value whereas futures anticipate a market value or price at an agreed delivery date in the future)
      1. <https://fred.stlouisfed.org/tags/series?t=oil>
      2. About Economic Research at the St. Louis Fed - The widely used FRED® data service is updated daily and allows 24/7 access to over 500,000 financial and economic data series from more than 85 public and proprietary sources.
3. Transformation of the Data (data cleaning and transformation required)
   1. Imported CSV’s into Jupyter Notebook
      1. The unemployment data was already annualized. (see Figure 1 in Section VI, Appendix)
      2. renamed columns
      3. deleted unwanted columns
      4. deleted unwanted rows i.e. foreign countries (only presenting US data)
      5. deleted data segregated by male and female retaining only total unemployment
   2. Crude oil data (see Figure 2 in Section VI, Appendix)
      1. renamed columns
      2. data was originally reported on a daily basis
         1. created a column containing only the year
         2. deleted the yyyy-mm-dd column
      3. dropped years less than 1955 to match the unemployment data
      4. groupedby year and calculated the yearly averages
4. Load (final database, tables/collections, and why this was chosen) (see Figure 3 in Section VI, Appendix)
   1. The gist of my interests in these datasets is that I have worked in oil & gas for approximately 20 years. For the first 15 years or so of my career oil prices were good to even great at times. So, from a professional perspective this became my normal or what was expected. About 4 years ago, the oil and gas market entered a depression and the recovery has been slow especially in the deepwater sector. With the most recent historic drop in oil prices I thought it would be interesting to see if there was a correlation between oil prices and unemployment on a national scale. However, I am sure that low oil prices are very attractive to industries such as logistics and trucking, commercial airlines, and to consumers in general.
5. Appendix

Figure 1



Figure 2



Figure 3

