

**ASSIGNMENT FRONT SHEET**

**Course Name: ALY6050 20906 Intro to Enterprise Analytics**

**Professor Name:**  **Dr. Christopher Johnson**

**Student Name: Dong Quoc Tuong (Lukas)**

**Student Class: Fall 2019 CPS Term: B. 2020**

|  |
| --- |
| **Module 2: Predictive Analysis**  **Completion Date:t March 3rd Due Time:12:00am** |

**Statement of Authorship**

*I confirm that this work is my own. Additionally, I confirm that no part of this coursework, except where clearly quoted and referenced, has been copied from material belonging to any other person e.g. from a book, handout, another student. I am aware that it is a breach of Northeastern University’s regulations to copy the work of another without clear acknowledgement and that attempting to do so renders me liable to disciplinary procedures. To this effect, I have uploaded my work onto Turnitin and have ensured that I have made any relevant corrections to my work prior to submission.*

**Tick here** to confirm that your paper version is identical to the version submitted through Turnitin

**Executive Summary**

Nowadays, business prediction results are essential for the evalutation of the future performance of any company. As the world becomes more connected, planning and prediction procedures establish a concrete foundation against future uncertainties. This paper addresses the question on how Predictive Analysis can influence Business in the Literature Review. Then, we will perform the Regression Analysis in the Energy dataset and have a look at the model’s application to the Health Care sector.

**Background and Literature**.

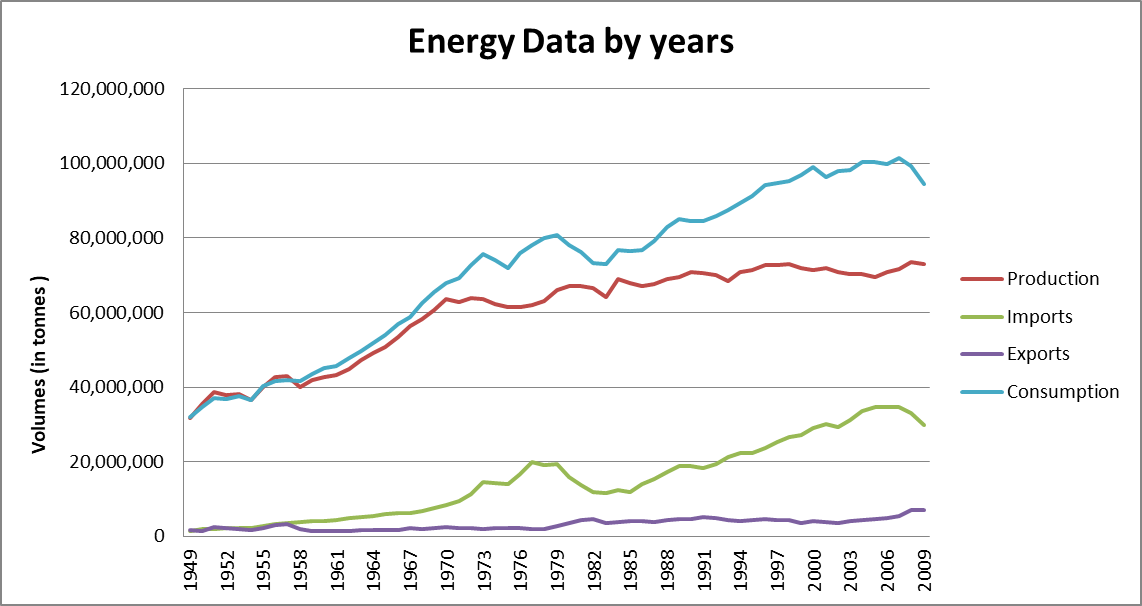
Those newly introduced to business forecasting might want to know the difference between Predictive analysis and Business Intelligence (BI). BI allows users to slice and dice data in order to address past or present events of a company: what happened, what is happening and why it happened. Thus, the BI results are often static reports or dashboard which could potentially inflexible. On the other hand, Predictive Analysis offers an insight to the future through estimated outcomes (targets) of interest. It not only is a deeper, more proactive method but also ignores the predefined cube data structure. In short, unfamiliar ones can think BI and Predictive Analysis is the two end of the analysis spectrum. (TDWI Research & Halper, 2014). Nevertheless, the process of reading the crystal ball has many drawbacks. They are technological (Required data’s volume, difficult source data’s access; difficult results implementation and difficult integration into IT systems) and organizational (lack of awareness, in-house experts’ shortage and results’ low accuracy). Attaran & Attaran (2018) have proposed a conceptual framework for successful implementation of analytics in different organization:

1. Executives involvement
2. Careful planning and scoping
3. Establish an exciting business case
4. Determining business process that needed to be corrected
5. Determining KPI
6. Choosing experienced workers
7. Providing appropriate tools for the analysis needs.
8. Maintaining a solid data governance program
9. Granting analysts access to useful data.
10. Moving on to the next analytics model if fail

At the end of the day, it is important that business leaders to are aware of how a predictive analytic initiative will impact external and internal users, their products and multiple stakeholders before choosing it.

**Analysis**

We will analyze the Energy Production and Consumption dataset that provides the energy information on the production, consumption, imports and exports. The dataset is from1949 to 2009. Overall, America suffers from slight trade deficit with exports volume is equal to that of imports volumes at the start. Over the years, the gaps between these two figures widened as America began importing more goods rather than exporting to satisfy its massive growing consumption rate.

****

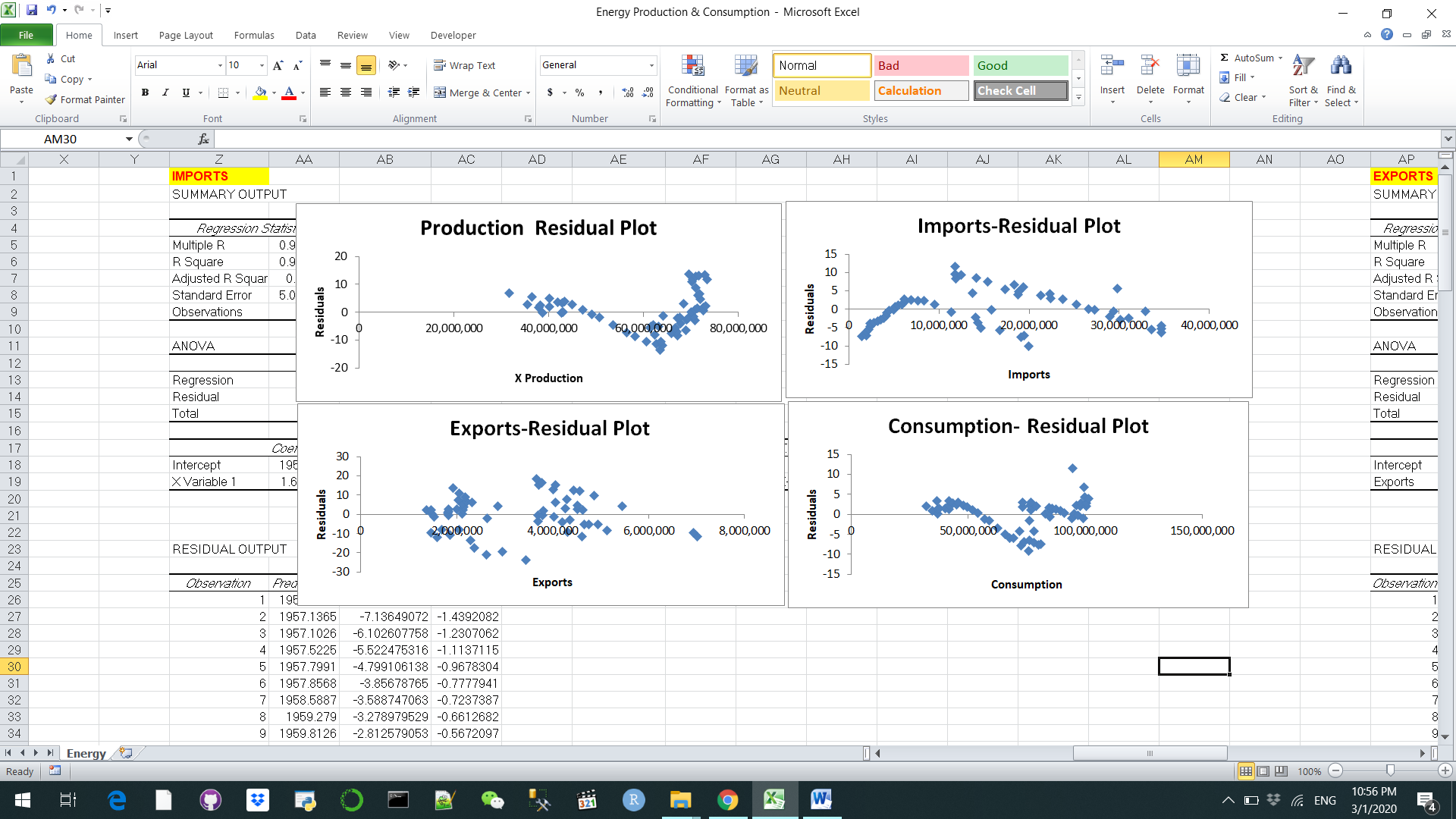
From a respective point of view, almost all variables have strong positive correlation with time with consumption being the strongest. Exports, though shared some strong correlation, remained the lowest out of all variables. The same patterns when we look at the R Square and the adjusted R squared, which detailed how close the data are to the fitted regression line. We can rest assured that the model explains approximately more than 70% of our variability of the response data around their means. Therefore it is no surprising when exports’ standard error is the highest because standard error indicates how wrong the regression model is on average in predicting the results. In another word, the lower R Squared and Adjusted R Squared are the higher standard error is.



Residual Sum of Squares is the total variation in the dependent variable that was ignored by our Regression model and it is Exports that has the highest number. If we set up the null hypothesis that Years and Variables are not related with each other, then we definitely favor the alternative hypothesis because all of the Significance F are noticeably lower than 0.05, particularly Consumption variable.



When plotting the Residual plots for all three variables, we see that all of the points on Imports, Exports, and Consumption are randomly dispersed. Hence the linear Regression model is appropriate for the data. On the other hand, Production Residual plot resembled a U shape so we would have to go for a non-linear model. In conclusion, we will use Imports, Exports, and Consumptions variables will be used to develop the forecast and justify further investigation



**Application.**

The health care industry has been applying Predictive analytics to process gargantuan amounts of information so as to provide lifesaving diagnoses or treatment options before anything bad happens. Walgreens is one of the trailblazing companies in providing such personalized cutting edge service. But I will have a look at a much more intimate issue at hand: Scheduling beds for operated patients in the Emergency department. It is very hard for the hospital bed managers to accurately identify the days of the week where the department stumble access block because of the pre-assigned inpatients beds for the elective surgery. With the current bed management approach, one scenario often happens is the full ED on Monday morning and by midday, the hospital has been put on ambulance bypass. In order to address such issues, they just cancel elective surgery patients last minute, but this strategy frustrates not only administrative and medical staffs but also patients. Boyle and his partners (2008) tried to tested multiple Regression Models to in an effort to manage inpatient beds prospectively. The data was collected from Emergency Department of two hospitals over five years. So the researchers built a Regression Model to manage inpatient beds prospectively, allowing them to book elective surgery on specific times and days where there is low demand from the Emergency department. Predicted outcomes from the Regression Models were compared with observed administration data over a period of 6 months. The most accurate method was linear regression with 11 dummy variables to model the monthly variation (MAPE=1.79%). Such method has the lowest mean absolute percentage, bearing in mind that this does not necessarily correlate to the model with the greatest R2 value when fitted to the training data. If we use the simple averaging technique, it shows similar forecasting performance to the Regression Analysis along with the finer time periods.

**References**

Attaran, M., & Attaran, S. (2018). Opportunities and Challenges of Implementing Predictive Analytics for Competitive Advantage. *International Journal of Business Intelligence Research*, *9*(2), 1–26. <https://doi.org/10.4018/ijbir.2018070101>

Boyle, J., Wallis, M., Jessup, M., Crilly, J., Lind, J., Miller, P., & Fitzgerald, G. (2008). Regression forecasting of patient admission data. *2008 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. https://doi.org/10.1109/iembs.2008.4650041.

TDWI Research, & Halper, F. (2014). *Predictive Analytics for Business Advantage*. Retrieved from <http://www.sapevents.edgesuite.net/previewhub/ITAnalyticsContentHubANZ/downloadasset.2014-03-mar-17-21.predictive-analytics-for-business-advantage-pdf.pdf>