**Future In Tech: Data Analytics Project - Note to Students**

**The Business Scenario**

Sunshine Deserts has been in operation since 1986, and the CEO is reviewing the company performance. There are a number of areas she wishes to consider as part of that review and to assist her, she has hired a data analyst.

Her main aim is to undertake a retrospective analysis of the company between the last full operating year 2001 and the year the business was established, 1986. Once the analysis is complete, the requirement is to present the findings to the board. **S**he would like to capture information regarding each of the following with commentary from the DA to position the analysis

Firstly, some top level figures regarding the business

* The total number of employees, the total salary spend and the gender balance between in 1986 and now in the year 2001 – across the company as a whole and then within the various departments.

**Secondly,** she is interested to compare salary spend in the year the business was established (1986) versus the spend in 2001

* How has the salary profile changed within the company, 1986 vs 2001 across each of the departments?
* Is there a metric that you would recommend as a basis to compare the spend within each department? Why would you consider one Vs another?
* Given the gap between the two periods of time 1986 and 2001, what should be taken into account when comparing the figures? Would you recommend a mechanism to adjust the figures?

As you analyse the salary spend, are there other analyses you would recommend that would inform the CEO on the various changes taking place? Is the data limiting your analysis in anyway and how might you overcome those limitations?

**Thirdly,** gender balance in 2001 Vs 1986:

* What is the gender profile across the departments?
* How could these be explained?
* What about salary? Is there a salary bias?
* Are some departments more gender biased than others? Has this balance changed?

The board will want to understand trends you identify, are some departments more balanced than others? Why might that be the case?

**Finally,** how about employee churn in 2001 (only)

* What does employee churn look like when you consider the various departments. Note the “to\_date” in the dept\_emp indicates that an employee has left that department e.g. 9999-01-01 indicates the employee is still within the department.

Churn = #employees leaving in given yr / total employees in the dept in given yr

* What conclusions can you draw?
* In 2001, are particular departments more susceptible to churn

What conclusions can do then draw from the analysis? Once again consider the recommendations that you would make to the CEO. Are particular department more susceptible to churn?

Presentation of Results:

* At this point you will have some high aggregated data regarding the company, gender balance, department specific data related to employee salaries and churn.
* You must now consider how best to present this data within the Tableau application. What options can you consider to put the data into a format which lends itself to presenting the information within Tableau?
* You should consider an approach which takes into account the ease of making changes to the data. CEOs change their mind and therefore you want to make life easy for yourself. You will want to automate the process as much as possible to avoid continuous rework within Tableau

Remember you are presenting to a business audience. **Their time is precious and they demand is actionable information. Facts based on evidence and associated recommendations or conclusion.** You will get 10 mins to present your core findings, thus no more than 4-5 slides. They must be impactful and tell the story you have identified.

**Building the Database and Loading the Data.**

You have been supplied with 4 datasets:

* **Salaries,** detailing the payments to each employee
* **Employees**, detailing personal details and hire-date
* **Dept\_emp**, detailing where each employee works and how long they have worked their – note 9999-01-01 indicates they are still working in that department
* **Departments**, which links a department identifier to the formal name of each on

Before beginning your analysis you will be required to review the dataset, note their content and structure and then create an data model – an Entity Relationship. When building the model, you should consider the data types and the allocation of primary and foreign keys.

Once you have created the model, you should then load the data from the datasets provided and begin your analysis. Be sure to check the integrity of the data once loaded to ensure that all data is transferred to your various tables.