**Summary of Challenges face by Worker’s Compensation Claims Management**

**1.Process**

The standardized process followed by the management can be sectioned in three main steps.

The first section may be: When a worker of a participating company is injured in performance of her/his job, she/he can file a claim for the medical expenses and loss of wages related to the injury.

The second section may be: When a claim is filed for the first time, a unique claim identifier is created and the details of the claims are recorded.

The third section may be: The claim details are updated as payments are made to settle the claim. Payments can be made for various activities related to a claim, such as doctor visit, pharmacy, etc.

**key challenges**

Make decisions using data prevents organization management teams from gut feeling and erroneous decisions. However, dealing with big data is not an easy task because of many challenges it presents, ranging from data collection step to reliable and accurate decision-making stage. Between these two steps, Businesses working with big data face consistently many and various data management challenges and our business study case is not exempted.

The Business research case shows that the workers’ compensation claims management faced many key challenges as far as it works with big data as well. Following the process defined by the company in the compensation claims management, we can point out at least seven key challenges.

At least three key challenges can be identified at the first section of the process which is: When a worker of a participating company is injured in performance of her/his job, she/he can file a claim for the medical expenses and loss of wages related to the injury.

**The first key challenge faced by the management is the Data availability**

Since, the management relies on the data filed by claimants and recorded to make decisions, the availability of data becomes a challenge. The volume of the available data depends then on how much data is filed and how much of filed data has been recorded, resulting in a challenge the management need to handle.

**The second key challenge resulting from this first stage of the process is obviously the quality of the data**

The data quality refers to the reliability of the data. For the data to be useful to the management team, it must meet some specific criteria. The data must to be complete, unique, valid and consistent to be useful.

**Complete:** The management team need to assure that the data present is a large percentage of the total amount of data needed. Missing data can make the available data useless.

**Unique:** For the datasets to be free of redundant or extraneous entries, the management team needs to make the data unique by creating a unique ID for each claim recorded.

**Valid:** The data collected from the claimants are mixed in first place. It does not conform to the syntax and structure defined by the business requirements. This becomes a key challenge since invalid data need to be transformed into format that meet the syntax and structure defined by the business to be useful.

**Consistent:** Every change or update made to the dataset need to be consistently represented in a standard way throughout the dataset. any database transaction must only change affected data in allowed ways. For example, if we were storing a number in a database, only the numerical values are allowed. when a payment is made to settle a claim only payment value needs to be affected.

**The third key challenge faced by the management is the Data Integrity:**

For the dataset to be useful, the quality only is not enough. The data integrity becomes a necessity. In addition to the quality, the management needs to assure that data are in context to be useful.

**Big Data Analytics**

**Role of external data**

**2.Report on the different types of data issues in the Dataset**

* Duplicate ID (Claim Identifier) in data Aggregated Transaction. that needs to be made unique
* The dataset is unsorted and needs to be sorted
* Missing values for variables Incident Date, Return to Work Date, Claimant Opened Date, Claimant Closed Date, Employer Notification Date, and Claimant Received Date in ClaimsDataFall2020 data.

**Plan of action to clean and prepare the dataset for subsequent analysis**

**Data Wrangling**

* Delete rows with duplicate ID (Claim Identifier) in data Aggregated Transaction.
* The dataset is unsorted and needs to be sorted
* Delete rows with missing value for values for variables Incident Date, Return to Work Date, Claimant Opened Date, Claimant Closed Date, Employer Notification Date, and Claimant Received Date in ClaimsDataFall2020 data.

**Data Visualization**

**Create Count of Merge Cleaned ClaimsDataFall2020.csv for each Claimant Type.**

**Count of Merge Cleaned ClaimsDataFall2020.csv broken down by Injury Nature.**

**Count of Merge Cleaned ClaimsDataFall2020.csv for each Received Date Year. Details are shown for Return to Work Date.**

**Sum of Indemnity Paid for each Claimant Closed Date Year.**

**Count of Merge Cleaned ClaimsDataFall2020.csv, Is Denied and Is Fatality.**

**Indemnity Paid, Other Paid, Total Paid, Total Recovery and Total Reserves broken down by Claimant Closed Date Year.**