Strucutre: problem—data collection—database design—report system—prediction

**KPI**

Osha KPI is a tool that can help you to monitor your department’s accident rate. I will include 3 of them in here today:

1. OSHA Recordable Incident Rate (ir)
2. Lost Time Case Rate (ltcr)
3. Severity Rate (sr)

OSHA has established specific mathematic calculations that enable any company to not only monitor their own performance, but also they are comparable across any department, division or group. Those KPIs are pretty self-explanatory. Let’s see an example:

This OSHA KPI chart is tracing its 5 years trend. As we can see the Incident Rate is down trend, but both Severity Rate and Lost work days rate is going up.

It’s my honor to have you guys here to talk about data science and data-driven decision-making process. Everybody, nowadays, is talking about data-driven decision-making, but how？Without a clear understanding about data science project, the data-driven decision-making is just a dream.

In Risk Management, our cloud database store all accident related information, for instance, I can offer you the top costly division in your department, top accident location, cause for the accident and how much it was to settle the case, things like that. Risk management is storing data properly in different level of detail, with a carefully designed cloud database, and automation report system.

Today I will talk about the standard process in data science project to help you make the data driven decision making process come true.

There are 3 roles in data science world. First Data Engineer, who build data pipelines to stream data across platforms so that data from different department are comparable to each other. Based on the data infrastructure created by engineer, data analyst then design database, reporting system, and define KPI. Last but not least data scientist who apply cutting-edge predictive modeling, such as regression, DecisionTree and Naural network.

Personally, I am the combination of data engineer, data analyst and data scientist.

Basically, Data science process consist of:

problem statement

**Step1: Problem Statement**

The first step for a data science project is always Problem Statement: Have you ever tried to write a problem statement? You might write one at the beginning of a design project, during discovery, to explain what you’re working on. Here are some examples I’ve encountered:

* How much PPE we should buy keep an appropriate inventory level.
* How can we combine our records with data from another dept.
* We want to reduce the accident rate how can we do that.

These examples are useful in their ability to provoke conversation, but they hardly from the basis for a detailed design project. Problem statements fall apart when they’re too vague or too prescriptive. Ultimately, you and your team need to judge whether a problem statement is working for you. Does it help your team recognize what their focus is? Does it keep you aligned in the same direction?

Data collection , Database Design, report system development

**Data collection & Database Design**

Data **Collection** is the hardest part about building report system, and predictive modeling. it’s in collecting quality data and identifying relevant features. So that we need to be goal-oriented during data collection, in other word, keep our problem statement in mind. Data collection is also crucial in database design. When we are talking about collection, Level of details is the first thing come to mind. For example If we didn’t store client’s birthday information, we will not be able to tell his age when doing analysis and report.

We use **database** to store data where data is used to reflect the real world. So, I also call database as mini-world. Often, stored data is used for reporting, analysis and transformation. It protects our information from leaking, and because of the database schema, workers can have a better understanding about the blueprint.

Let’s say if you need a solid data collection method to get start with data driven management, or you want to cooperate with other department and need a data transformation solution, or you want your team to be more data-oriented by developing a database to deal with specific high risk daily tasks, feel free to contact with Risk management and we are willing to help. Our mission is to help you lower the rate of accident.

Risk Management is managing accident data for all departments. But because of that, our database may lack of deeper level of details. Because it’s for everybody. In this situation, our insight could still be too general sometimes. However, if you have your database which keep record about daily operation details, just like the JHA report Gabe introduced earlier, then I can data mining deeper and root cause will be much clearer.

Take Animal Care and Control for example, they do have their database to record animal’s information. We can then do the analysis and tell them the Risk Level for each breed, different season, etc. So that they can be aware of danger.

Predictive modeling

**Predictive modeling**

Predictive model is one step further. After you have a solid data infrastructure, different algorithm could be applied to help with prediction, classification and clustering.

I March-2020, I joined covid-19 command center, the manager over there want me to help her with the daily report. However, they don’t have a clear data infrastructure. I mean, they pick the.

In summary, I can help you to