**Introduction： slide 1+2 🡺 2mins 慢慢说**

My name is Michael Jin. It’s my honor to have you guys here to talk about data science and the data-driven decision-making process. Everybody, nowadays, is talking about data-driven decision-making, but how？Without a clear understanding of data science projects and a solid data infrastructure to support, the data-driven decision-making is still far away. Today I will introduce OSHA KPI, a.k.a Osha key performance indicator, which is crucial to track the accident history, then I will talk about the standard processes of data science project come with examples. Finally, I will talk about what Risk Management can offer and what we need to work on together.

Generally speaking, I am working on three aspects of the data science world:

1. I am Data Engineer who builds data pipelines to stream data across platforms so that data from the various departments are transferable and comparable.
2. Then, I am a data analyst. Based on the engineer's data infrastructure, I design the database structure, the reporting system and define the Key Performance Indicator to monitor the trend.
3. Finally, I am a data scientist who applies cutting-edge modeling for predicting, such as regression, decision tree, and Neural network.

In Risk Management, our cloud database stores all accident-related information. For instance, I can offer you the costly top 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 various levels of detail, with a carefully designed cloud database and automation report system.

**KPI: Slide 4—7 3mins**

To get the start, Osha KPI is a tool that can help you to monitor and measure how many incidents have occurred or how severe they were. OSHA has established specific mathematic calculations that enable any company to monitor their performance, but they are comparable across any department, division, or group.

I will include 3 KPIs today:

1. The first one, Severity Rate, is a mathematical calculation that describes the Number of lost days experienced compared to the Number of incidents experienced.
2. Secondly, OSHA Recordable Incident Rate. The is calculated by multiplying the Number of recordable cases by 200,000 and then dividing that Number by the Number of total labor hours at the company
3. Thirdly, the Lost Time Case Rate is a mathematical calculation that describes the Number of lost workdays per 100 full-time employees in any given time frame.

From the Osha KPI definition above, Osha KPI cares more about Loss workdays instead of the medical cost that makes OSHA KPI a great complement tool for regular accident report which focuses more on medical cost alone.

Overall, these KPIs are pretty self-explanatory. And now, Let’s see an example from one of our department:

I extracted two charts from the traditional reporting system. The one on the left is about Number of accidents by year, and the right one is about the average accident cost by year. Even though the Number of accidents is pretty much the same over the year, the average accident cost is down-trending, telling me that this department is getting better in accident control from the medical cost point of view.

This is the Osha KPI I talked about earlier. It is tracing five years accident trend. As we can see, the Incident Rate is a downtrend, but both Severity Rate and Lost workdays rate is going up. We can assume a lower accident rate, but at the same time, each accident is more severe from a lost workday point of view.

Kpi plays a critical role in helping managers monitor their crew more effectively and give a direction to dive in when strategy-making. One last thing I want to mention about Osha KPI is that it has been standardized so that managers can compare statistically significant KPIs with other departments or with their own past safety performance, even though the size of your team is changing. Osha KPIs may help managers to determine where their department may need additional program assistance to prevent accidents from happening again. And managers can learn from other department accident prevention policy by comparing KPI ratio.

**Data Project: 2 mins**

Do you guys have any question about the Osha KPI?

Well, I believe you understand Osha kpi better, and please send us a request if you want Osha KPI and other accident related report.

Now Let’s talk about data project. Typically, there are 4 stage during a data science project.

They are Problem Statement, Database design, Reporting system, and predictive modeling.

**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, for example:

* How much product 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 3mins**

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. What’s more, level of detail will affect the performance of a predictive model.

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 to develop report, analysis and transformation. It protects our information from leaking, and because of the database schema, showed on the left, managers can have a better understanding about the blueprint of a problem.

**Reporting System**

With a solid database, we can then develop automated **Reporting System**. Instead of working full day to find data everywhere, reporting system can help with decreasing Latency, increasing accuracy and consistency, reduce human error and easy to share.

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 in advance.

**Predictive modeling**

The core of Predictive model is algorithm. Think algorithm as a smart kid, and data as the book that the kid is reading. Better book or better data quality can help us train a more reliable algorithm to help with decision making.

**Covid-19 center 2mins**

With that being said, here is a sample project That I have done for Covid-19 command center.

In March-2020, I joined covid-19 command center, our mission was to give out free PPE to help non-profit facilities get over the hard time. I created the data-collection-form to collect PPE request data, streamed information into one database and then developed central data warehouse to connect all covid-19 related database in one place. In the central data warehouse, we can easily find customer’s PPE burn rate, inventory level from different PPE warehouse, PPE giveout policy, price, budget, etc. With all of those information grouped together, we can then understand the covid-19 situation more accurate and faster, and we were able to predict how much PPE we need to buy in order to keep our inventory at a appreciate level.

This is one chart Command center is using to track the total PPE give out. One surprising fact in this chart is from November 2020 to January 2021, the total PPE give out is showing an exponential pattern, which is an alarm for us and force us to make new PPE purchasing prediction.

We use both machine learning algorithm and statistical model to forecast the PPE burn rate.

Six-sigma is the final prediction method that we decision to follow. Six-sigma is a technique that widely accepted in process improvement. For example, Airbus is using six sigma standard to produce flight, and the statistically accident rate of a flight produced under six sigma standard, is 700 years.

In this prediction, I give out PPE burn rate from 1 sigma to 6 sigma, so that the management team have a few choice to decide how much PPE we should really buy, depending on the budget.

**Summary 1min**

Today we started with 3 Osha KPIs, Severity rate, Incident Rate and Lost workday rate. They are useful tools for you to monitor accident situation. And unlike the monthly report that mainly focusing on the medical cost, the OSHA KPI care more about the impact of Loss workdays.

Then we talked about data science project. Just like what I have done in the covid-19 command center, We can help you to develop your own data platform so that data mining and prediction can go into a deeper level.