**Demo Script**

**Opening Slide:**

Good afternoon, my name is Emily Rechan and I am a technical specialist here at IBM. I specifically focus on business analytics with a concentration in business analytics tools. Rui Fan also is on the call today, she acts as a data scientist at IBM and specializes in Data Science Experience. We are both very excited to have the opportunity to speak with you and appreciate you taking the time to listen in.

Today, we want to share with you how to gain more from your data and how to reach a higher level of competitive insight.

**Slide 2:**

In order to do so, we will first review your company goals, second I will share the overview of the solution with you, and finally conclude with a demonstration. After, we can discuss any questions not addressed during the demo, but please feel free to interject at any point. This will ultimately assist us in creating next steps and moving forward together.

To start off, lets first discuss your company’s current environment and any concerns you may have.

As the final decision maker at your company, you mentioned ….

**Slide 3:**

Your business analysts and data scientists are having a difficult time collaborating together as they work on completely separate platforms and the data across the organization is siloed. This is ultimately affecting your bottom line because your unable to effectively rate your customers risk level when it comes to changing or creating existing and new car insurance coverage.

The two main aspects you want to concentrate on in a new platform are:

* Ease of use for your business analysts
* Obtaining deep insights for your data scientists

Each of these aspects will allow for your company to better analyze specific customers, while creating better car insurance policies. In today’s demo, I will take on the role of the business analyst while Rui acts as the data scientist

Now, If there are any other concepts you would like us to expand on please let me know as we go through the demo.

**Slide 4:**

To begin I think it is important to note the process we are going to see here today. This includes taking our data and consolidating it. Thinking back to your company's current issues this will directly solve your concern with having siloed data, as everyone who has access to this information can locate it from one tool.

Once the data is all in one place, a connection can be made to multiple other tools. In this case we are going to use two that can directly benefit your company; however, there is no limit and services can be used interchangeably.

To start off our demo, Rui will share with you our data consolidation process. Rui…

**Data Connection – DB2 warehouse on cloud:**

Thank you, Emily. Data consolidation is critical in order for every member of your organization, with access, to have the ability to locate necessary data from one place. This will enhance your organizations efficiency by allowing an agile flow of information. This can all be done through cloud services.

While thinking about the different job titles and responsibilities in your organization you may find that these different personas may modify the exact same data set in different ways. In this portion of the solution we will show how collaboration between each persona can be done seamlessly by having access to a universal cloud based data warehouse for your company.

Today you will store and manage vehicle collision data and make connections to corresponding analytics services for analysis.

Home page: As you can see here, this is the home page. On the left hand side there are various functionalities. While here you can see three of the main capabilities, including loading data/working with tables/ and running SQL.

For this demo, we will first deploy the load data function. [click load data] You can see that there are many options for loading data. This can be done from different hubs, off your personal device in a local file, or other cloud services. One thing I really want to specify here is that IBM partners with twitter, which provides you the opportunity to use twitter data in analysis. One way this can be beneficial is through sentiment analysis.

For the sake of this demo however, we will load data directly from our desktop.

After we choose the corresponding data file, it is loaded in this data base. This will automatically generate a preview of the data being uploaded into the database. For the purpose of this demo we will create a new table and load it.

If you want to change the name of a table that can be done in seconds (change name to Vehicle Collision – don’t have to say it)

And now you are finished.

As you can see, in just a few seconds we were able to take various data sets and place them in one platform. This provides every position, whether a business analyst or data scientist, access to the data they need to effectively add value to the company.

Connection: Now that our data has been added into a database allowing open collaboration between different departments, we want to start analyzing and generating conclusions. This can be done through various services.

It is important to recognize that connections can be created very quickly. (have the connection already done – go to the connection page with all the different services) Here you can see a few of the services that can be connected.

To summarize, we were able to create a space to store our data creating a platform for all employees, with access, to enter and use in their analysis. This is crucial as business analysts and data scientists will now have insights that complement each other while remaining consistent.

Emily, will now show you the insights she has found while taking over the role of the business analyst.

**Data Analyst – Watson Analytics:**

Thank you for making that connection Rui. Within this business analytics service, you will have insights at your fingertips. This service will speak directly to your main goals: ease of use for your business analysts, collaboration for your employees, and assist in making more accurate decisions through knowledgeable insights.

Before we get started, I’d First like to reiterate that this service also adds to the benefit of collaboration. As you can see there is a similar data connections area that shows various services to connect with. This will allow each department to build off one another’s findings.

Now, when we go to our homepage you will see three sections Data, Discover, and Display

The data tab is where you will find your data sets, as you can see our vehicle Collison database data is automatically uploaded from the connection that Rui made in the data consolidation phase. The main areas to note here are that you can access both shared data sets, which will allow your employees to build more unique and creative insights together, as well as, personal data sets that each employee can use for their own work. Also, you will find that your analysts will be able to shorten the amount of time they spend on refining data as they can do so from this platform.

Next we will move to the discover tab, this is where you can navigate through your data set to find insight into your business problems. Here you have the ability to ask your data questions in natural human language. This means that our platform is able to actually understand our questions and find relevant answers for us, this will ultimately save your analysts time and in turn provides yourself critical information to make decisions.

Another valuable aspect are our starting points. These are suggestions that the platform automatically offers you as it has already processed your data. This means before your analysts put in any input, they already have a starting point. So, let’s try it out.

As an auto insurance company the first thing you may want to know is… what vehicle type has the most collisions? This service will automatically create starting points associated with our question, so let’s choose the first one here. …. As you can see a word cloud is created for us, and passenger vehicles are clearly the most likely to be in collisions.

Now as a business analyst you may want to change the visualization to be a little clearer or just in a different format, that can be done with just a few clicks. Again, here you’re offered suggestions as the platform already recognizes the various visualizations that will work best with these criteria. Let’s choose the bar graph. Now, the data is a bit more organized. However, we still are seeing some irrelevant data such as bicycles having no collisions, which is unnecessary for us to pay attention to for auto insurance. We can easily filter it out in a moment. As you can see your business analysts no longer have to spend hours editing and refining, everything can be done right here and instantaneous.

Before we move on to the overall insights that I was able to find in relation to your company, let’s quickly summarize.

In the data tab you were able to locate and refine your data

While in the discover tab you were given the opportunity to ask that data questions, edit visualizations, and filter those visualizations in the matter of minutes.

Both of these areas speak to your goal of ease of use.

Now, I recognize that shared insights are crucial for your company to be able to operate effectively. This can be easily done through the display tab. This offers the option to create a dashboard, infographic, or expert storyboard. Dashboards are great for exploring your insights and creating an interactive display. Each visualization will change as you begin filtering and adjusting the visualizations. However, for the sake of this demo, as the business analyst, I have already built a dashboard and a storyboard with some crucial insights.

Storyboarding allows you to control the narrative and allow your audience to easily pay attention to your main findings. Here you are also able to continue asking your data questions and can save your discoveries in a more unique and understandable way, which then can be shared with colleagues and key decision makers.

This here is our storyboard home, the main areas to concentrate on here is that you can customize your board and even ask further questions to your data just as I showed in the discovery phase.

When looking at our top findings, you will see here two dashboards that have already been filtered to represent the main areas we need to focus on.

Our annotations are shown by this dot. This is beneficial for when storyboards are being shared throughout the organization and someone wants to quickly look and understand the visualizations being shown. This will enhance communication between viewers and collaboration.

The dashboard here tells us that there have been 13,867 injuries caused by collisions, passenger vehicles are the most common cause of collisions, and this is typically due to driver inattention and driver distraction.

This information had me interested in knowing three main areas: where are these collisions happening, specifically what streets in that area, and does time of year affect these collisions?

Our next dashboard was able to explore these questions….

Here we find Brooklyn has the highest number of collisions, so I filtered our data to show what streets specifically in Brooklyn had the most collisions which are Atlantic avenue and Flatbush avenue. Lastly I ask if the time of year made any changes, and we find that in march and April there is a spike in collisions.

You may be thinking now, well how does this help us? Well, for instance, when deciding risk levels for current or potential customers, you can send this information to your direct associates who will now incorporate where individuals live, travel to often, or vehicle they drive into their risk assessment. For example, an individual that lives in Staten Island, but travels each day down Atlantic Avenue in Brooklyn for work in a passenger vehicle may be given a higher risk score with much higher rates, this will assist in reaching your goal of identifying and decreasing high risk customers.

[Slide to DEMO slide]

There are still many more way to dig deeper into this data set and more insights to be found, but this information creates a good foundation for a business analyst to share with their colleagues. This can easily be done through the shared folder.

As the business analysts, I now want to know even more about this data, so I am going to send over my findings to our data scientists, Rui Fan.

**Data Scientist - Data Science Experience:**

Thanks, Emily, for introducing what we can do as a data analyst. From Emily’s analysis we were able to find that the most common vehicle for collisions are passenger vehicles. In this section. As the data scientist I will now enter the platform and find additional information important for your company. With this platform, we are able to build different visualizations and prediction models, and data scientists are able to collaborate with each other to finish projects.

IBM is now a data science company and its reputation is widely recognized by many business partners. This service is one of the cloud solutions provided for data scientists, not only for work, but also for collaborating with their colleagues and getting involved in different data science communities. There are many functionalities like notebooks, tutorials and even data sources. With this platform, your data scientists are able to work on a same project, sharing their notebook with R/Python/Scala scripts which is run on the newest spark service, while also sharing ideas. Even your analysts who doesn’t know programming can participant to do data science work, because you can import code from open source communities to do business.

Let me show you how to use this platform.

*Show:*

As you can see, this is the home page. There are four sections: the menu bar shows all of functionalities; the ‘Recently posted’ shows some resources in a data science community; and the ‘updated projects’ shows the projects you have recently used; In the end, these are some helpful links for better use of this platform.

In order to connect this platform to the data warehouse we have, we must first make a connection. This is similar to the connection we previously made between the warehouse and the business analysts service.

*[Go to ‘Data Service’ – ‘Connections’ – create new – Data Service to connection to the dashDB I created just now.]* Another benefit is that there are multiple external data sources you can leverage, including Amazon redshift and Microsoft Azure. This means you do not necessarily have to lock in with only IBM solutions but can utilize the services you already have in place as well.

To complete the connection, you will want to choose the database associated with our data set.

*[Then I choose this one, and choose a database, BLUDE]*

Now that our connection has been made, what can I as the data scientist do with the vehicle collision data? Well, I can create a new project or use an existing one. For the sake of this demo, we will use one I have already created. – specify the name of this project, choose the spark service and built-in object storage you want to work on and then ‘create’. This is the dashboard of a collaborated project. You will see why I call this a ‘collaborate’ later in this demo. Then I would like to go back to ‘Overview’ and create a new Jupiter notebook for further analysis. In this window, the thing I really like is that even you are not a code lover, you still can directly import notebook/R, python script in this notebook by past a target URL. In this way, everyone is able to do data science work regardless of knowing program as long as they carry the knowledge of a data science. [Go to the pre-built notebook] Here is the notebook I created for this use case. In this window, if you want to insert a certain data, directly click here: ‘Insert to code – insert to R data frame’ to make your life easier. After you have all of your data ready in the notebook, then you can start doing advanced analytics. First is data manipulation, which will lead to several important graphs:

1. This is a histogram of Number of collisions per person in each borough. This shows the possibility that people who live in these boroughs may get in a car accident. From the visual we can see that people who live in Manhattan have the highest possibility of a dangerous collisions, while the people in Staten Island are less likely. These results vary from what we saw in Emily’s analysis as we are now looking at each individual person, not just total collisions. This platform allows us to narrow down our analysis even further then before. This can also be driven from the next line graph.
2. This line graph shows the pattern of accidents in each area over time. In 2015 we see 247 accidents happened compared to the decline that we see in the first month of this year, around 150 accidents happened. This does indicate things are getting better over time. I also want to note that these findings match those that Emily found earlier as in march 2015 there was a spike in collisions specifically in the Brooklyn borough. We were able to build off of that information to find an overall trend.
3. After we collect the line graph, we assume that the number of collisions is related to time of year. From the following graph we can see *[click],* that assumption is verified. In the beginning months of the year, we have the highest number of collisions in all of the boroughs, while Brooklyn has still had the highest collision with close to 12000.
4. Now to dive deeper into the time of the year, we will focus on the number of accidents happening each day of the week. With still Brooklyn as the highest and if you go out on the weekend, there is a less amount of collisions. As a data scientist you may assume this is due to the lack of work week traffic.
5. This visualization indicates the reasons for a vehicle collision. And Driver distraction appears the most in accidents happening in NYC, confirming Emily’s analysis earlier. .

So why is this data important to your company? Well, one way to decrease the number of high risk customers, is through education. You can share this data with your current customers to make them more aware of collision factors. For future customers you could potentially create a survey to fill out, which would ask questions regarding their daily travel habits. This would ultimately assist your associates in identifying high risk customers. Another way you can enhance this analysis is in relation to the earlier months of the year. You may ask, is weather a factor? Possibly due to high snow fall, high winds, or ice during this time of year? This could be a possibility, having your data scientists add weather data would allow for further analysis.

It may also be beneficial for your company to have a predictive analysis of how many accidents will happen in the future. To do so, I have divided the data into training data and testing data to build the prediction model. As a result, the testing error rate is 0.9. Actually I investigated several models before building this, and linear regression seems to fit the this better and it is much simpler and flexible.

In addition to the analysis, I would like to use shiny to build a nicer and interactive map to show collision pattern. Rstudio is leveraged in this situation. *[Tools – Rstudio]*:

1. Just in one easy click, a shiny app is shown. This is so flexible for data scientist because they can edit the scrip to meet whatever visual requirement. Here is just one UI possibility for us to check.
2. There are two menu bars, one is an interactive map to help you have a better understanding where and what kind of collisions happened – by selecting different conditions, the corresponding place of accident is shown in this NYC map. Also, you can investigate the reasons for those collision. And the other can help you drill down to see what really happened in each collision. In this interface, it still possible to filter by different conditions.

As a data scientist I am interested in adding more data sources in this project, for example, to add weather data to investigate the relationship between bad weather and collision, and number of tourists’ data to evaluate the most dangerous street and number of people walking through. This will allow further analysis.

Let’s go back to this notebook to explore other valuable areas of this platform. [click back to the project window]

To address collaboration again, as I see this is one of your main goals to reach, we will us this small person icon, which allows us to add collaborators to this project, and one the left hand side, every contributor is able to leave their comment and share ideas. [click community]

By this icon, you can add addition notebooks/data assets/models in this project to allow deeper learning. ‘Community’, gives you the opportunity to check articles that stay on top of big data, share your newest ideas to a specific algorithm, or even publish your most satisfied notebooks. If you have any questions regarding the topics of data science, you can find the answers here.

*Summary:* As you can see, this is a fast and easy way to drive insight, and in the meanwhile, you are able to learn information from many sources and share projects with colleagues. As you can recall, it is integrated, powerful, collaborative and so informative. This platform speaks to your main goals of deep insights as we were able to build on Emily’s original insights, collaboration, and discover ideas to decrease the overall high risk customer.

**Closing:**

Thank you Rui, that was very informative.

Well, it seems as though we have come close to the end of our time here today. However, I do want to quickly review the three services we have covered.

1. Data Consolidation
   1. A one stop shop for your company’s data sets that allows each department in your organization to access. This service also creates connections to various other services, making it much easier for business analysts, data scientists, and other roles to complete further analysis.
2. Business Analysts Service
   1. This platform allows for easy collaboration through shared files
   2. Refining steps to clean your data sets,
   3. Discovery capabilities through natural human language
   4. Data presentation options through various displays such as story boarding

Ultimately this service creates an easy, creative, and collaborative area for analysts to solve problems

1. And lastly, our Data Scientist Service
   1. Which makes anyone a data scientist as you don’t have to be a programmer or an engineer, this interface is user friendly for all experience levels
   2. Also, we were able to find deeper insights into our collision data allowing us to have concrete information to decrease high risk customers.
   3. This platform is collaborative and provides various resources to excel as a data scientist.

Each of these services have been able to speak to each of your business goals and initiatives. We want to thank you for taking the time to listen in today and please feel free to reach out with any questions or concerns.

Our contact information can be located here on this slide, please feel free to reach out at any time. Thank you and hope to hear from you soon.