|  |  |  |  |
| --- | --- | --- | --- |
| Data Science Experience Demo | | |  |
| Hortonworks + IBM Briefing Sessions 2017  This is a suggested timeline to provide a high-level demo of Data Science Experience or DSX. *Reminder: this is a marketing event and there is 15 minutes of time allotted for the demo so keep it high-level.* | | |  |
| *Start* | *End* | *Task* | *Duration* |
| *0:00* | *1:30* | **Welcome audience, introduce yourself and provide brief introduction to Data Science Experience or DSX**   * What is DSX? – it promotes data science as a team sport. Projects, data, models can be shared securely and easily across all members of the team. * Why use DSX? – it enables end to end data science; from data wrangling, model development, through deployment. | *1 min*  *30 secs* |
| *1:30* | *2:30* | **Brief overview of the DSX landing screen**     * Recently published – mention these are from the Community which include Articles, Data Sets, Notebooks and Tutorials * Recently updated projects – mention these are the projects * Navigation – hover over Projects and Docs | *1 min*  *30 secs* |
| *2:30* | *6:00* | **Discuss the Project overview page**     * Notebooks – mention the most recent Notebooks for the project will appear here (5 max) and the options and statuses   + View all – navigates to the Analytics assets tab where all the projects Notebooks will appear   + View a Notebook – by clicking on the Notebook name   + Edit a Notebook – by clicking the pencil   + Shared – if the Notebook is shared the  icon will appear   + Running – if the Notebook is running the  icon will appear   + Locked – if the Notebook is being edited the  icon will appear   + Actions – brief over the available actions Publish on GitHub, Publish as Gist, Share, Duplicate, Delete * Data assets – mention the most recent Data assets for the project will appear here (5 Max) and the options   + View all – navigates to the Analytics assets tab where all the projects Notebooks will appear   + New data asset – expands a right-panel with options to add a File or Connection; same as clicking the  icon * Bookmarks – mention items discovered in the Community area can be bookmarked  to this project and will appear here (5 max)   + View all – navigates to the Bookmarks tab where all the projects Bookmarks will appear   + Explore community – navigates to the Community page; same as clicking the top-nav Community link or clicking the  icon | *3 min*  *30 secs* |
| *6:00* | *7:30* | **Discuss the Collaborators tab**     * Collaborators – mention the Project’s collaborators will appear here and the options   + Change permissions – dropdown list to quickly change a collaborator’s role   + Remove collaborator – delete a collaborator from the current project   + Add collaborator – invite collaborators to the current project and set their permissions | *1 min*  *30 secs* |
| *7:30* | *9:30* | **Discuss features of a Notebook** – click on the  to Edit the Part 1: Data Exploration - Python (public) Notebook     * Note – Point out this Notebook’s Kernel is running Python 2 with Spark 2.1 * Change kernel – click on the Kernel menu item and hover over the Change kernel submenu to show the available combinations for the Notebook * Show Markdown – click the 1st cell and mention it is configured as Markdown * Show Code – click the 3rd cell and mention it is configured as Code * Add, Cut, Copy and Paste Cell – mention the menu items to Add , Cut , Copy  and Paste  a Cell * Move a Cell – mention menu items to move cells up  and down | *2 min* |
| *9:30* | *10:30* | **Show Notebook Comments** – use the Part 1: Data Exploration - Python (public) Notebook     * Show collaboration – click on the Comments  icon to open the Comment panel. Mention real-time commenting at the Notebook level * Mention notifications – hover over the Notifications  icon | *1 min* |
| *10:30* | *11:30* | **Add Data to the Notebook** – use the Part 1: Data Exploration - Python (public) Notebook     * Insert a Data Connection – click on the Find and Add Data  icon to open the Data panel. Click on the arrow next to TEST\_SUM.csv to expand the options. Click on *Insert SparkSession DataFrame* to add code to the new cell | *1 min* |
| *11:30* | *13:00* | **Explain the Banking Churn Notebook part 1** – scroll through the Part 1: Data Exploration - Python (public) Notebook     * Explain Churn\_Rate.csv – briefly mention the data file contains 3 years of churn rate broken out by quarter * Mention Data Visualization – briefly mention you can easily import powerful visualization libraries such as PixieDust, Brunel and Maplotlib to view your data in Tables, Charts, Graphs and Maps * Explain Test\_Sum.csv – briefly mention the data file contains US customer profile with a true/false column whether they have churned * Aggregate by State – briefly mention the grouping STATE and INCOME data to visualize on a Brunel Map customer income by state * Distribution of Churn – briefly mention the data is being analyzed for possible answers to the increase in churn * Click to Part 2 – click on the link to quickly jump to the Part 2: Deploy Churn Model - Python (public) Notebook | *1 min*  *30 secs* |
| *13:00* | *14:30* | **Explain the Banking Churn Notebook part 2** – scroll through the Part 2: Deploy Churn Model - Python (public) Notebook     * Brief overview – this notebook will utilize TEST\_SUM.csv to create a model to train, test and validate the data for future identification of customers that may churn. * Mention steps   + Load the data – load the identified fields from TEST\_SUM.csv into a Data Frame   + Split the data – randomly split the data into 3 Data Frames to train, test and validate the model   + Train the model – create a pipeline for the model with the train data   + Test the model – get the Accuracy and Test Error score by running the test data through the model   + Create and plot a ROC curve – visualize the true and false positive rate of the model with a ROC (Receiver Operating Characteristic) curve   + Deploy the model – deploy the model to Watson Machine Learning and test the new API end point by running through a test profile and receiving a score. Briefly mention this API end point is accessible by other applications | *1 min*  *30 secs* |
| *14:30* | *15:00* | **QUICK QUESTIONS???** | *30 secs* |