Clyde Icuspit: Welcome to the IBM Code Tech Talk: Making data science simple. This talk is being recorded and a replay of the presentation will be available for viewing on our YouTube channel immediately after the call. https://developer.ibm.com/code/videos/tech-talk-replay-making-data-science-simple/

From Luis Moraes: Greetings to all IBMers from Sao Paolo, Brazil
From PAOLO ARCELLA: is possible have presentation file (slides) at the end of the call?
From Kathy Ghaneei: Charts are posted to the above link
From Steve Martinelli: @Paolo, the slides will definitely be posted in the link Clyde just shared!
From PAOLO ARCELLA: thank you so much
From Steve Martinelli: long live house stark!
From Florian Lutz: :D
From Mike Prentice: FYI, if you do not watch this GoT, you completely miss the point being made
From Jeff Reichard: David mentioned that there is a shortage of Data Science "Experts". What skills and knowledge does a Data Science expert have? Is it experience with certain tools? platforms?
From Jeff viezel: GOT and Friends - I love it!
From Leo Zuniga: Audio is gone..right?
From Steve Martinelli: @Leo, seems OK to me
From Jeff Reichard: Can just anybody say that they are a data science expert?
From Leo Zuniga: my bad, local issue thanks!
From peter neirynck: Will the presentation (in ppt format) made available internal IBM for reuse?
From Kathy Ghaneei: Peter we can only post pdf to the site bc of WordPress. if you email me I can send you ppt.
From Mark Sturdevant: A lot of people can do a little data science, but most donâ€™t have all the math/stats/data-science training to be that true data science expert
From Mark Sturdevant: tools help fill that gap
From Steve Martinelli: Well said Mark
From pascal brenner: What is your email Kathy?
From Kathy Ghaneei: kghaneei@us.ibm.com
From Phil S: datascience.ibm.com allows a free trial signup/ is there an internal (w3) version which IBMers can use for our own internal teams?
From Mark Sturdevant: Iâ€™d recommend just using the free trial
From vickie dorris: does ML include optimization algorithms ie: like ilog Cplex
From Kevin Minerley: A "data scientist" almost always needs a subject matter expert to help steer the avenues of research
From pascal brenner: thanks kathy
From Jeff Reichard: What is the difference between Data Science and Machine Learning as IBM sees it.
From Ramesh Menon: How do we audit a DL / ML model - Details? This is important to eliminate human bias.
From Phil S: k. iâ€™d like to better manage our current sources of data and itâ€™s not a short-term objective. when does the billing meter start with the free trial?
From piero proietti: @Kat may I suggest to upload the PPT in BOX and share the link?
From Thomas McLaughlin: is "Data Engineer" a career path?
From Steve Martinelli: @Phil bluemix has a chargeback model for internal billing. i'll try to see if i can find the info we used
From Steve Martinelli: @Thomas we'll see if David can give him opinion on that question :)
From Ryan Mayor: https://console.bluemix.net/catalog/services/machine-learning
From R K: Is this session recorded for later use?
From Kathy Ghaneei: @RK, yes it is recorded. it will be on Youtube.
From Peter Mark: lost screen share
From R K: is this code handwritten?
From Steve Martinelli: @RK - yes, all tech talks are recorded
From Kathy Ghaneei: https://www.youtube.com/playlist?list=PLpryjkO3KF2w4Q68Hsi-O2i3WsbAZ-7Z3
From Phil S: thanks @Steve
From Vickie Dorris: we don't see charts
From Nam Duong: sharing?
From Vickie Dorris: only Scott
From Peter Mark: lost screen share
From Joe Kozhaya: are you sharing?
From Conrad Prukop: change to share screen
From Devesh Pant: can someone provide the urls that are being spoken about?
From Mamnoon Jamil: We don't see anything
From Burton Boucher: not seeing
From Stefan Clausen: All we see is you, Scott
From Steve Martinelli: fixed :) From Steve Martinelli: the github repo is here: https://github.com/IBM/pixiedust-traffic-analysis
From Carlos Fabricio Fernandes: So, is it possible to use current mobile data to map city traffic like a doppler ultrasound analyses. _x000B_Also, we could map one particular IMEI and build its proximity network and map terrorists cells, right?
From Rishiraj Deb: is it a tool like Microsoft's Power BI?
From Mark Sturdevant: the python is running in the cloud. You have spark and object storage if you need it From Steve Martinelli: @Deepak when you select a notebook to run, you also specify a Spark service to run it on. Bluemix offers various levels of Spark services, there is a free tier too From Aditya Shirode: Isn't the legend covering the rest of the graph?
From Deepak C Shetty: Ok, but here what Scott is showing, is there spark involved, its plain simple python code, so is spark in the picture, is my Q?
From Carlos Sanchez Rivas: I would like to know how can we create our own database? by using non structured data and structured one….can I do this here?
From Les Andrzejewski: Do you use Cognos on the back end?
From Tommy Burke: is there any information on the chargeback for these services for IBM internal users?
From R K: is this code hand written?
From Rishiraj Deb: It seems like a visualisation tool, like Microsoft's Power BI

From David C Martin: shouldn't™t that analysis compare the actual values and calculate level of deviation from the norm
From Mark Sturdevant: For a database it is really easy to connect to cloudant NoSQL from a notebook From David C Martin: shouldn't™t you control for seasonality, eg week of year?
From Rajendra Shankarappa: How do we load our own data?
From Mark Sturdevant: I usually load data by adding it to object storage

From Steve Martinelli: @Rajendra, sure, we use data that was freely available, see "accidents = pixiedust.sampleData("https://data.sfgov.org/api/views/vv57-2fgy/rows.csv?accessType=DOWNLOAD")" in the notebook
From Mark Sturdevant: There are some simple ways in the ddx menu to drag-and-drop file
From Mark Sturdevant: but also API for object storage
From Stefan Clausen: Is there going to be a recording made available of this session?
From Mayur Raja: Do you know what format the data is in and what fields you can select on?
From Mark Sturdevant: There are other options if your data is accessible via some API (DB or URL, etc.)
From Steve Martinelli: @David there's only one season in SF though ;)
From Kathy Ghaneei: Stefan, yes. replay will be on youtube: https://www.youtube.com/playlist?list=PLpryjkO3Kf2w4Q68Hsi-O2t3WsbaZ-7Z3
From Joyce Li: Does the pixie interactive chart still work if the notebook is in read-only mode?
From R K: Is it possible to put the Recording and Slides on the BOX?
From Steve Martinelli: @David, the notebook is not a complete solution, we hope that it encourages people to take a look at data science with a fun example
From Joyce Li: It would be helpful if we can share the interactive chart using the share capability in DSX
From Steve Martinelli: so seasonality was not taken into account :)
From alex kotlov: how can we integrate the pixiedust maps into bluemix apps? any APIs?
From Carlos Fabrecio Fernandes: We could also map mobile usage by provider, identify shadow areas and show which operators could best invest in one particular area to install RBS. x000B. Also, using this information, have a photo of the area just prior a natural catastrophe and guess how many people could be under debris.
From David C Martin: this is aæœdata miningaæœ not aæœdata scienceaæœ. where is the hypothesis?
From Deepak C Shetty: PixeDust Vs Jupyter - when to use what? any quick insights
From Deepak C Shetty:?
From Kevin Minerley: Source of "free" info although it may be throttled these days: https://catalog.data.gov/dataset
From Mark Sturdevant: @David not sure if there is a hypothesis/conclusion in this one, but a DSX notebook is a good tool to add that. This demo does focus more on the explore and display part
From Mayur Raja: Can this be used to import SMF data to analyse the performance of IBM subsystems?
From Deepak C Shetty: Can we create enterprise/production level apps using pixiedust/jupyter + DSX or its only for making DA learning fun?
From Mayur Raja: Why do we call it a notebook?
From Mayur Raja: What is ZMQ?
From Mark Sturdevant: @Mayur good question. As you edit the text cells along w/ the code cells you get the feel of a notebook where you state the problem try some code, document conclusiona€½ Take notes to explain what you did and what you found
From Steve Martinelli: @Mayur you can find more info about jupyter notebooks here: http://jupyter-notebook.readthedocs.io/en/stable/notebook.html
From Mayur Raja: @Mark and Steve, thank you
From Patricia Lorenzo Cruzado: All this tools work in different platforms and with different tools or you need a specific setup for them to work?
From Savi Abraham: Can there be a Playaround Walk Through set of instruction shared where in one can try their hands on creating 1 of the 3 samples embedded. In this way one gets a hang of all the steps and sequencnes that needs to be completed to get an out put version. And then give an option in the sample code where in one can add code lines and see what it does to the final output etc
From Steve Martinelli: @Mayur, i missed the ZMQ reference, but that's probably a reference to zeromq, a popular messaging library
From Mark Sturdevant: @Patricia DSX is hosting jupiter notebooks for you so you can run from anywhere
From Mayur Raja: @Steve, thanks, I wanted to know if it was IBM MQ for z/OS
From Tommy Burke: how does DSX compare to https://datascientistworkbench.com/?
From Ali Khanafar: Would be cool if you can publish a prompt along with the chart...
From Simon Laws: How are the shared results secured?
From David Lambert: The "share URL" uses an IP address ... what's the lifetime on it?
From David C Martin: and the underlying data? as CSV? as JSON?
From Marina greenstein: What is different in your set up that you have SHARE button?
From Robin Starr: This is exciting!!! I gather that a chart for the Bay Area can be created for rush hour traffic?

From Christoph Vischer: Charts are static, right? How to do a refresh and load updated data?

From Patricia Lorenzo Cruzado: If you share the results on the web how do you secure that information?

From Mark Sturdevant: @Tommy I haven’t tried PixieDust outside of DSX though

From David C Martin: and pixie dust is just one choice? what about Brunel?

From Mark Sturdevant: @Christopher you can always go back to earlier cells that load/refresh the data and then run cells down to the pixie dust cell

From Angshuman Roy: Where is that IP address hosted? Is it part of setup? Can we have a URL with domain name instead of IP address?

From Kathy Ghaneei to Mark Sturdevant(Privately): Mark can u look at youtube questions pls

From Alex Kotlov: any plans/dates when the security/authorization will be integrated?

From Mark Sturdevant to Kathy Ghaneei(Privately): trying to. not much to answer there right now

From Lucian Paucean: Short Q: I see this as a complex tool for charting/visualization of data ... the question is if and where the Watson learning layer is applied ... how is this different of google maps stuff that is already out there ... where is the value added ove the competition

From Kathy Ghaneei to Mark Sturdevant(Privately): ok thx

From Luis Moraes: This Wheater App URL is there too?

From Marina Greenstein: Can you please share the url of what he is talking about right now?

From Devesh Pant: url for the blog post pls

From Steve Martinelli: David blogs a lot on medium, check out his posts at https://medium.com/@DTAIEB55

From Angie Krackeler: great session, thank you

From Tony Hon: is pixiedust an open source?


From Mark Sturdevant: @Tony yes it is open source

From Patricia Lorenzo Cruzado: Thank you for sharing the url

From Srinivas JVS: PixieDust can be installed locally on windows 10? Or do I need to use DSX for trying out PixieDust?

From Robin Starr: This is exciting!!! I gather that a chart for the Bay Area can be created for rush hour traffic?

From Frederic Petit: How does it compare to Polly and Tableau?

From Frederic Petit: Plotly

From Patrick Dantressangle: How scalable is pixie dust? It is a js framework that probably has difficulties to display hundred of thousands data points: correct?

From Tapas Som: Just think about if you did not have the Powe Point features to do charts - think about Portlets - tookss hundreds of hours. If you know the nature of data (say time series) and want standard analytics to be visualized - these are must have oob features

From Dean Wilson: How do you compare and contrast this with Cognos Analytics and Watson Analytics?

From Daniel Turco: Is the presentation going to be shared with us?

From Mark Sturdevant: @Patrick I think you would use python data frames to get your data prepared before you give it to pixie dust

From Tapas Som: SaaleForce.com built these tings around 2010 oob and kicked IBM portal out of an accoun tthat I was involved in

From Amit Shah: What is difference between IBM Data Catalog and Alation Data Catalog

From Frederic Petit: @patrick What about pandas frame?

From Jeff viezel: Can you interface this with Cognos Analytics?

From Kevin Minerley: Wolfram’s CDF format is similar to PixieDust

From Patrick Dantressangle: it is not about that. it is about the displaying capabilities. basic Js frameworks are not very scalable for many thousands datapoints

From Patrick Dantressangle: how far did you tets this (in the display? not the data frame which is a given)
From Mamnoon Jamil: Statistics/Operation Research/ Machine Learning expertise are needed to be Data Scientist
From Altaf Kagalwalla: what's the difference between Quants and data scientists?
From Gopal Indurkhya: https://www.virtualitics.com/ provides excellent Virtual/Augmented Reality for data scientist
From Gopal Indurkhya: do we have something like that?
From Jeff viezel: A data scientist is someone who gets away from opinion and â€œgut feelâ€ to cut through the BS and deliver fact based analysis
From Leo Zuniga: Thanks a lot!
From Leo Zuniga: from Mexico!
From Luis Moraes: Great Presentation,, Thanks guys!
From Scott DAngelo: Thanks Everyone!
From PAOLO ARCELLA: thank you so much for all