Hands-on Workshop

Justin Broxton, IBM
Agenda

• What is Watson Analytics? (10 minutes)
• Demo (15 minutes)
• Workshop (85 minutes)
  – Data Format (5 minutes)
  – Logging in (5 minutes)
  – Navigation (5 minutes)
  – Explore (20 minutes)
  – Refine (10 minutes)
  – Predict (20 minutes)
  – Assemble (20 minutes)
What is Watson Analytics

Quick start intuitive interface

Natural language dialogue

Key business driver insights

Dashboard and storytelling authoring

Easy data upload and Refinement capabilities

Explore
Ask questions and interact with the results to discover patterns and relationships that impact your business.

Predict
Find predictive insights hidden in your data. Learn what drives each behavior and outcome - and take your next steps with confidence.

Assemble
Create interactive and engaging dashboards and infographics and tell persuasive stories to share and communicate with others.

Refine
Enrich and tune your data to discover patterns and get new perspectives on your business.
Refine and **personalize** data as needed

Look at data metrics for each data field. Create calculations, groups, and hierarchies.

Change column titles, data types, modify default aggregations, sort data order.
Analyze and quickly find root causes in your data
Uncover hidden patterns in your data
Build engaging dashboards

Sales by Product
- Part Group
  - Car Accessories
  - Car Electronics
  - Performance & Non-auto

Win or Loss Filter
- Loss
- Won

Total Opportunity Value
- 5,286,743,512

Average Days in Sales Cycle
- 7.9
Workshop
Getting logged in

1. Go to: www.watsonanalytics.com
2. Click “Sign In” in the top right corner
3. Enter your username and password provided

Result: Everyone should see a screen that looks like this:
Basic Navigation

Click on the title in the top bar to see a list of open items.

Next, click “Welcome” to get back to the home screen.

Click to close an item.
Basic Navigation
Part 1:
Loading Data
pages 13-24 in guide
Data Format - Good example
Data Format – Incorrect format

<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
<th>G</th>
<th>H</th>
<th>I</th>
<th>J</th>
<th>K</th>
<th>L</th>
<th>M</th>
<th>N</th>
<th>C</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sales Report</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2015 - Units sold</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Products</td>
<td>Region</td>
<td>Jan</td>
<td>Feb</td>
<td>Mar</td>
<td>Apr</td>
<td>May</td>
<td>Jun</td>
<td>Jul</td>
<td>Aug</td>
<td>Sep</td>
<td>Oct</td>
<td>Nov</td>
<td>Dec</td>
</tr>
<tr>
<td>4</td>
<td>Product A</td>
<td>NA</td>
<td>273</td>
<td>412</td>
<td>369</td>
<td>293</td>
<td>268</td>
<td>223</td>
<td>308</td>
<td>419</td>
<td>124</td>
<td>484</td>
<td>303</td>
<td>158</td>
</tr>
<tr>
<td>5</td>
<td>Product A</td>
<td>Europe</td>
<td>448</td>
<td>429</td>
<td>414</td>
<td>350</td>
<td>454</td>
<td>438</td>
<td>417</td>
<td>176</td>
<td>420</td>
<td>224</td>
<td>417</td>
<td>240</td>
</tr>
<tr>
<td>6</td>
<td>Product A</td>
<td>AP</td>
<td>262</td>
<td>277</td>
<td>364</td>
<td>439</td>
<td>375</td>
<td>445</td>
<td>375</td>
<td>374</td>
<td>138</td>
<td>274</td>
<td>447</td>
<td>126</td>
</tr>
<tr>
<td>7</td>
<td>Product B</td>
<td>NA</td>
<td>205</td>
<td>116</td>
<td>146</td>
<td>134</td>
<td>172</td>
<td>108</td>
<td>334</td>
<td>287</td>
<td>129</td>
<td>270</td>
<td>421</td>
<td>355</td>
</tr>
<tr>
<td>8</td>
<td>Product B</td>
<td>Europe</td>
<td>281</td>
<td>382</td>
<td>130</td>
<td>309</td>
<td>233</td>
<td>179</td>
<td>345</td>
<td>382</td>
<td>427</td>
<td>247</td>
<td>478</td>
<td>161</td>
</tr>
<tr>
<td>9</td>
<td>Product B</td>
<td>AP</td>
<td>464</td>
<td>269</td>
<td>419</td>
<td>244</td>
<td>374</td>
<td>238</td>
<td>241</td>
<td>419</td>
<td>141</td>
<td>463</td>
<td>389</td>
<td>500</td>
</tr>
<tr>
<td>10</td>
<td>Product C</td>
<td>NA</td>
<td>313</td>
<td>359</td>
<td>168</td>
<td>256</td>
<td>383</td>
<td>260</td>
<td>127</td>
<td>379</td>
<td>365</td>
<td>464</td>
<td>322</td>
<td>302</td>
</tr>
<tr>
<td>11</td>
<td>Product C</td>
<td>Europe</td>
<td>281</td>
<td>475</td>
<td>389</td>
<td>469</td>
<td>493</td>
<td>246</td>
<td>472</td>
<td>103</td>
<td>176</td>
<td>395</td>
<td>393</td>
<td>453</td>
</tr>
<tr>
<td>12</td>
<td>Product C</td>
<td>AP</td>
<td>433</td>
<td>127</td>
<td>223</td>
<td>233</td>
<td>419</td>
<td>486</td>
<td>226</td>
<td>386</td>
<td>426</td>
<td>377</td>
<td>110</td>
<td>200</td>
</tr>
<tr>
<td>13</td>
<td>Totals</td>
<td></td>
<td>2960</td>
<td>2846</td>
<td>2622</td>
<td>2727</td>
<td>3171</td>
<td>2623</td>
<td>2845</td>
<td>2925</td>
<td>2346</td>
<td>3198</td>
<td>3280</td>
<td>2495</td>
</tr>
<tr>
<td>14</td>
<td>Average</td>
<td></td>
<td>328.89</td>
<td>316.22</td>
<td>291.33</td>
<td>303.00</td>
<td>352.33</td>
<td>291.44</td>
<td>316.11</td>
<td>325.00</td>
<td>260.67</td>
<td>355.33</td>
<td>364.44</td>
<td>277.22</td>
</tr>
</tbody>
</table>
Data Format – Better, but not ideal

| A   | B         | C | D | E | F | G | H | I | J | K | L | M | N |
|-----|-----------|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1   | Products  | Region | Jan | Feb | Mar | Apr | May | Jun | Jul | Aug | Sep | Oct | Nov | Dec |
| 2   | Product A | NA    | 273 | 412 | 369 | 293 | 268 | 223 | 308 | 419 | 124 | 484 | 303 | 158 |
| 3   | Product A | Europe | 448 | 429 | 414 | 350 | 454 | 438 | 417 | 176 | 420 | 224 | 417 | 240 |
| 4   | Product A | AP     | 262 | 277 | 364 | 439 | 375 | 445 | 375 | 374 | 138 | 274 | 447 | 126 |
| 5   | Product B | NA    | 205 | 116 | 146 | 134 | 172 | 108 | 334 | 287 | 129 | 270 | 421 | 355 |
| 6   | Product B | Europe | 281 | 382 | 130 | 309 | 233 | 179 | 345 | 382 | 427 | 247 | 478 | 161 |
| 7   | Product B | AP     | 464 | 269 | 419 | 244 | 374 | 238 | 241 | 419 | 141 | 463 | 389 | 500 |
| 8   | Product C | NA    | 313 | 359 | 168 | 256 | 383 | 260 | 127 | 379 | 365 | 464 | 322 | 302 |
| 9   | Product C | Europe | 281 | 475 | 389 | 469 | 493 | 246 | 472 | 103 | 176 | 395 | 393 | 453 |
| 10  | Product C | AP     | 433 | 127 | 223 | 233 | 419 | 486 | 226 | 386 | 426 | 377 | 110 | 200 |
Data Format – Ideal

• Single column for measure
• Tools exist to transpose crosstabs
Data preparation

Available for Watson Analytics Q2 2016

• Intuitive data prep built for business users – no scripting or technical skills required
• Fastest way to analysis-ready data with point-and-click functions to easily manipulate, enrich and combine
• Access 30+ major databases, automatic extraction and preparation from multi-structured web pages, PDFs, text reports, HTML, JSON

© IBM 2016

ibm.box.com/waworkshop
Loading data

![Image of IBM Watson Analytics interface showing steps for loading data: 1. Click on "Add" and 2. Upload data.]

© IBM 2016
Loading data from a database

Where is the data you want to upload?

Select a source

- Browse
- Twitter
- IBM Cognos Report
- Dropbox
- Box

Or select a connection created for you

<table>
<thead>
<tr>
<th>Type</th>
<th>Name</th>
<th>Description</th>
<th>Modified</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>TestDashDB</td>
<td></td>
<td>Feb 3, 2016</td>
</tr>
</tbody>
</table>

Join Data

Choose how to join the data by selecting join keys. You can also select which columns to include from table ORDERDETAILS.

A. ORDER_HEADER

ORDER_NUMBER

10 of 10

B. ORDERDETAILS

ORDER_NUMBER

9 of 9

- A + matching rows
- B + matching rows
- All rows

Matches are case sensitive

Cancel Join
Loading data from a database

Available Data Connectors

- Amazon Redshift
- Apache Hive
- Cloudera Impala
- Cognos BI
- IBM dashDB
- IBM DB2
- IBM Informix
- IBM PureData for Analytics (Netezza)
- IBM SQL Database
- Microsoft Azure
- Microsoft SQL Server
- MySQL
- Oracle
- Pivotal Greenplum
- PostgreSQL
- PostgreSQL (on Compose)
- Salesforce.com
- Sybase
- Sybase IQ
Loading data from #Twitter

Where is the data you want to upload?

Select a source

- Twitter
- IBM Cognos Report
- Dropbox
- Box

Or select a connection created for you

Search

<table>
<thead>
<tr>
<th>Type</th>
<th>Name</th>
<th>Description</th>
<th>Modified</th>
</tr>
</thead>
<tbody>
<tr>
<td>Da</td>
<td>TestDashDB</td>
<td></td>
<td>Feb 3, 2016</td>
</tr>
</tbody>
</table>

Select the Twitter data you want

Enter up to 10 hashtags separated by spaces, for example: #ibmWatson #analytics.

#feelthebern #trump2016 #marcomortum

Enter dates and times in UTC. Your current time zone is UTC -4 hours.

- Start date (UTC): 2016-03-09 00:00
- End date (UTC): 2016-03-16 23:59

Data set name (200 character limit): #feelthebern #trump2016 #marcomortum

Hide estimates

Number of tweets available: 207650  Size (MB): 13.5

Your subscription level limits your data set to a random selection of 50000 of the available number of tweets. To make the data set more precise, change your parameters. To accept the selection, create the data set.

Upload data
Loading data from ‘flat files’ (xls, xlsx, csv)

1. Add
2. Upload data
3. Select a source: Browse

Where is the data you want to upload?

Or select a connection created for you:

<table>
<thead>
<tr>
<th>Type</th>
<th>Name</th>
<th>Description</th>
<th>Modified</th>
</tr>
</thead>
<tbody>
<tr>
<td>Da</td>
<td>TestDashDB</td>
<td></td>
<td>Feb 3, 2016</td>
</tr>
</tbody>
</table>
Dataset options

Click these dots to reveal dataset options (shown right)

Use this to replace your data set with a new flat file when the data changes/is updated.

If this data set came from a database, this would refresh the data from the database.
Your Turn:

Loading Data  (5 mins)

pages 13-24 in guide
Part 2:
Explore
pages 25-40 in guide
Getting started with Explore

Step 1: Click on a dataset (orange tile)

Step 2: Ask a question/search or choose a ‘starting point’
Having trouble asking a question?

Click ‘How to ask a question’ to get guided prompts that will help you form a question.
Explore – Working with visualizations

Everything in red can be clicked on

Satisfaction

Airline Name
Explore – Accessing Chart Data Options

Click the gear icon to see options for this chart.

Be sure to click X to see the chart legend again.
Sharing outside of Watson Analytics

You can share from Explore, Predict, or Assemble

This is the share button

ibm.box.com/waworkshop
Explore – Moving between pages

- Each Exploration can have more than one page
- Each visualization is its own page
- Use the page selector at the bottom to change pages

See all pages at once

Click on a dot to switch to a particular page

Create a new visualization/page
Explore – Local and Global Filtering

Local filters affect the current visualization (current page)

Global filters affect all visualizations (all pages)
Explore – Summarize

• Right-click on a visualization’s axis title to change how data is summarized (average, sum, count, etc.)
• You can also right-click on the column name in the chart’s data options (behind the legend)
Your Turn:

Explore (20 mins)

Pages 25-40 in guide
Part 3:
Refine
pages 41-46 in guide
Getting started with Refine

Step 1: Click on a dataset (orange tile)

Step 2: Click ‘Refine’
Refining Data

- Any changes you make to the dataset in Watson Analytics are saved in a refinement—the original dataset is unchanged.
- Refinement tiles are the same color as dataset tiles.
- Refine is used to augment a dataset with:
  - Calculations
  - Groupings
  - Hierarchies
- Defining subsets of data by filtering
- Renaming columns
Refine – Showing hidden columns

• “Rows” (Special column name) is hidden by default. This is essentially the row number, which can be useful if your data doesn’t have an index or unique column

• Automatically detected hierarchies are hidden by default
Refine – Filtering data

• Click on a column heading to filter the data by the values in that column
Your Turn:
Refine (10 mins)
pages 41-46 in guide
Part 4: Predict
pages 47-56 in guide
Predict

• Use predictive modeling to identify **patterns** in historical data that can explain **outcomes**
  - Which fields (inputs) are relevant in explaining a target outcome?
  - What combination of fields (inputs) lead to different levels of the target outcome
• Identify other interesting relationships in the data (un-related to the target)
What type of data works for the predict capability?

- There must be one or more target columns—columns whose values are influenced or driven by the values of the other columns.
- Your data set should include as many inputs (factors) as possible.
- The wider the data set, the better.
Getting started with Predict

Step 1: Click on a dataset (orange tile)

Step 2: Click ‘Predict’
Getting started with Predict

Name the prediction

Select the columns that are targets/outcomes in the dataset. Selected targets are shown to the right.
Predict – Field properties

• In some cases, it may be necessary to exclude a field from analysis. This can be done from the field properties.

• Click the three lines icon at the top left, then choose ‘Field Properties’
Predict – Field properties

- Field properties shows each field’s distribution
- Select a column, then select “none” to exclude it from analysis
Your Turn:

**Predict** (20 mins)

pages 47-56 in guide
Part 5:
Assemble
pages 57-85 in guide
Assemble – Resize Points

Rotates current item—does not resize

Most of our customers don't have airline status...

Opens chart data properties—does not resize

Resizes item, but with a fixed width:height ratio
Assemble – Contextual Menu
Click on a visualization to open a contextual menu (shown)

Some of our airports have long delays at peak times.

- Save to collection
- Duplicate
- Delete
- Change title
- Switch visualization type
Assemble – Filtering

• Clicking on any chart element will filter your dashboard by that element
• To remove the filter, **click on a blank space somewhere else in the same visualization**
Assemble – Data Tray

• Click the handle in the bottom-center to toggle the data tray
Assemble – Bringing in collected items

- The collection stores items you’ve ‘collected’ in Explore and Predict
- The guide refers to this as ‘pinning’ (old name)

Click this icon in the upper-right corner to open the collection:

Drag and drop to add items to the dashboard
Assemble – Changing summary, filter, or sort

- Click on a visualization’s axis titles to change between sum/average/etc., sort the data, or filter the axis
Assemble – Filtering a dashboard

Global filters

Local filters

Columns
- Axis label
  - Airline Name
- Color by
  - Add a column
- Value
  - Satisfaction
- Value
  - Add a column

Local filters
- Age Range
Assemble – Styling and Properties
Your Turn:

Assemble (20 mins)

Pages 57-85 in guide