Data Visualization and Dashboard Design Requirements
For the Capstone Project

Introduction

Overview

CPI Card Group's global operations demand the need for consistent and accurate data for decision making. The business operates mostly in a single transactional enterprise resource planning (ERP) system; however, in certain circumstances other data sources arise, including within the Service Bureau, customer relationship management (CRM), and Hyperion. The business has the need to make business decisions based on the entire outlook of the business. To facilitate this change, data must be stored in a centralized location allowing users to access in a controlled fashion. The Business Intelligence project will create a centralized data warehouse to store all of the business' data in a less complex fashion. Making the data warehouse functional, Oracle's Business Intelligence Cloud Solution will access the data warehouse as needed by users, allowing for faster, more informed decisions. In addition to allowing users to access the data through the BI tool, a set of standard reports and dashboards will be created and published.

Business Reporting Needs

The main purpose of the data warehouse is to track and compare sales and costs for major dimensions across time periods. Sales should also be compared to invoiced amounts for major dimensions and time periods. Costs should be tracked by component for labor, machine, overhead, and material so that standard accounting measures can be computed such as gross margin, contribution ratio, and related ratios. In addition, planning performance should be evaluated by comparing sales to forecasts and costs to budgets.

KPIs for Job and Shipment Trends

- What are job revenue trends by location over time?
- What are sales agent productivity from leads to jobs over time?
- What are production trends for jobs (time to subjob production) for entities over time?
- What are shipment trends for jobs (time to shipment) for entities over time as compared to shipment promised dates?
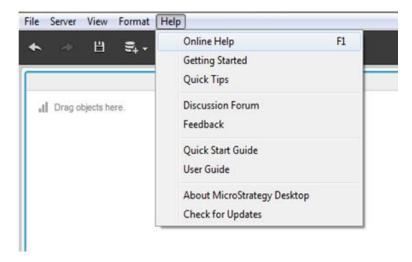
KPIs for Invoice Trends

- Which entities (such as customers, locations, and products) generate the highest invoice amounts over time?

- What are trends for invoicing of job amounts (time to invoice) for entities (locations and products) over time?

- What are trends over time for returns measured by the difference between invoice quantity and shipping quantity for products, machines, and locations?

KPIs for Financial Performance

- Determine the location and the machine which have the highest overall machine and labor cost. Also determine which location has the lowest budget overhead cost.

- Which location is seen to have higher forecast amount in comparison to the actual amount on the basis of time period?

Before Getting Started

For a quick reference or additional help within the reporting tool, you can click on help in the menu bar to troubleshoot, access video references or the user guide.

It is recommended to have a copy of the **User Guide** close by for a quick refresher of steps when needed.

Also review the video recordings available under the **Getting Started** > **See Video Tutorials**

Troubleshoot:

While doing your analysis, if you were to encounter a notification similar to the below, your only action is to close out of Desktop and restart your efforts from where you last saved your work. Throughout the document there are reminders to save your work progress, but please save frequently to ensure your rework will be limited.

Notification   ✕

(Object with ID "1DFF612E4E9E7FB9EC8C3C8927DAE890' and type 4 (Metric) is not found in metadata. It may have been deleted.)

OK

Getting Started

Note: Before creating visualizations please remember to derive your own dashboards and analysis. The below given charts are used to help you define metrics and attributes and should be used as a reference only.

Complete the project work in the following steps:

1. In the previous project you must have already connected MicroStrategy and the Oracle VM 12c database and loaded the required data in your database for this project.
2. Wrangle the data in MicroStrategy according to the format required as this is a raw data provided to you.
3. Create your own visualizations by referring the below dashboards and create your own dashboards to answer the questions given in each of the Work Sheet assignments.
4. Submit the .mstr file and your project summary by referring the below given dashboards.

What If I do not have MicroStrategy, Oracle VM and SQL Developer installed on my machine?

5. Use this document to install Oracle VM.

   Windows Virtual   Mac Virtual Machine
   Machine installation st installation steps.docx

6. Use this document to install MicroStrategy.

   Mac Microstrategy   Windows
   Installation Steps.docx Microstrategy Installat

7. Use this document to connect Oracle VM to MicroStrategy.

   Mac   Windows
   Machine-Connection t Machine-Connection t

8. Just a reminder: Connect the SQL Developer to the Oracle VM in order to load data, you might have already done this as these instructions for CREATE statements for the Production Data Warehouse are available in module 2 in a reading item that is locked until assignment 2 is submitted.

1. Once you have the tables ready in your database, connect and load it onto MicroStrategy.
2. Wrangle data in MicroStrategy in order to fetch the data for dashboard creation. This involves correcting data formats, you can use state abbreviations. Refer the below Date Wrangling Instructions for your reference.

Date Wrangling
Instructions.docx

⇨ The Data Wrangling tool creates a script of all the actions you perform on the dataset. This script needs to be saved (exported) as **<Student Name>_<Date>.txt** and included with your project submission at the end.

⇨ With the datasets added, take a moment to save the dashboard. When saving your dashboard, name it **<Student Name> - Capstone Project Analysis**

TIP   Data Blending allows you to combine data from multiple, diverse sources in a single dashboard and even visualization. This is particularly useful when you are trying to analyze data that is related but maintained separately and allows you to greatly expand your analysis.
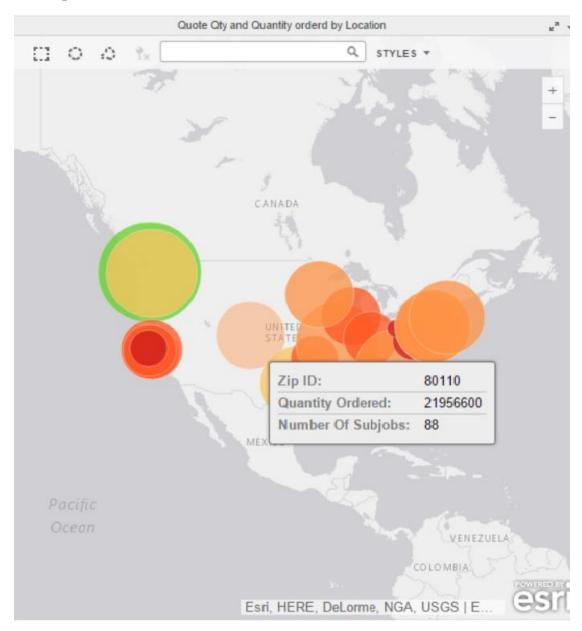
Please note: In this project, you will be creating 10 visualizations by using the Job, Subjob, Invoiceline, Shipment, Financial Summary Cost and Financial Summary Sales datasets. All these datasets should be linked to each other in order to check their relationships. For example, the jobid from W_Job_d table in the Job dataset should be linked to the jobid from W_Sub_Job_d table in the SubJob dataset. If they are not linked you will only see individual visualizations and not linked visualizations.

PART 1: Create four Visualizations Based on the KPIs for Dashboard 1 (D1):

Your first assignment is to create a presentation-ready dashboard to monitor the Key Performance Indicators (KPIs) for Job and Shipment Trends

[**Note:** Always follow 2 important steps for designing your dashboard, Wrangle the dates and Create multi-forms attribute for related attributes]

D1. Visualization 1: Quote Quantity and Quantity Ordered by Location (Top)

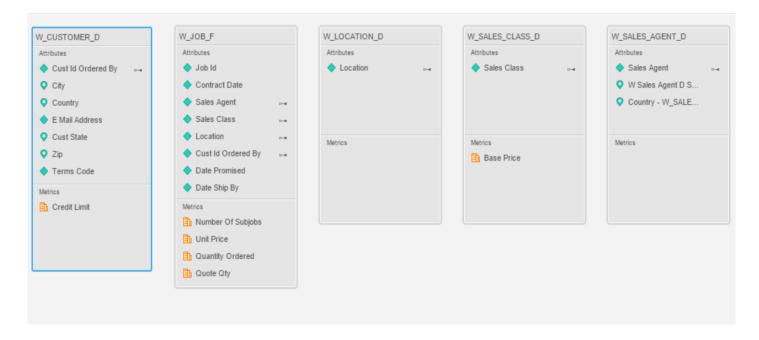**Source tables:** (Save this dataset as **Job** in the Dataset Panel)
Visualizations will be created on the basis of any dataset like Job, Lead, Sub Job, Shipment, INVOICELINE etc. So you can always select the attributes and metrics from the dataset for creating a visualization. Select the below tables to create this JOB dataset.

W_JOB_F
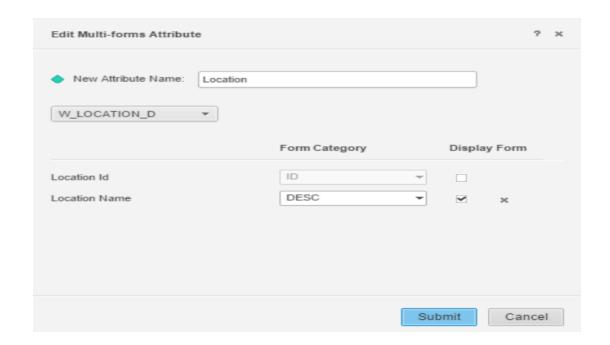W_Customer_D
W_Location_D

W_Sales_Class_D
W_Sales_Agent_D

Wrangle the dates in the table once loaded as given in the Data wrangling instructions, so that you display the correct time attributes.

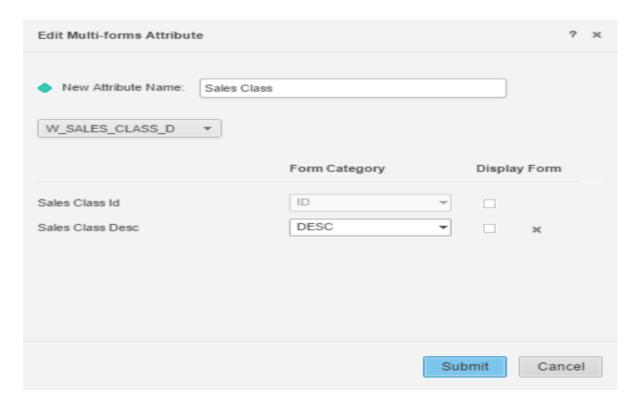**Mapping tables once loaded in MicroStrategy:**

- Select the mapping and the attributes according to the below given image. Just drag the attributes from one table onto another to map them. Once they are mapped you will be able to see a small sign besides your attribute

- Some metrics might not be seen in the Metrics section for each table, those are converted into attributes in some dashboards. So keep an eye on this.
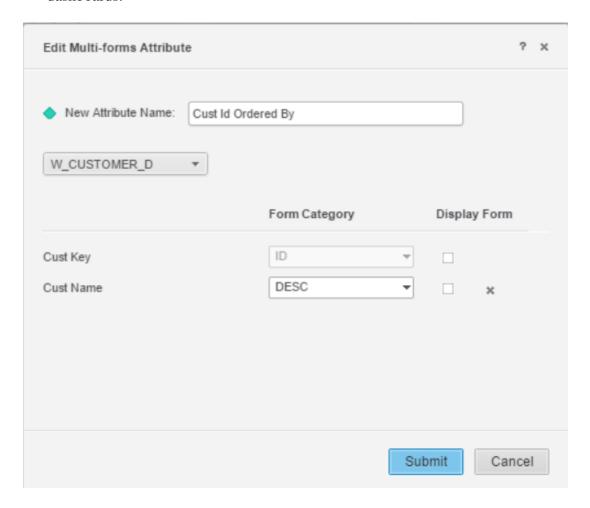

- Always combine Location ID and Location Name by clicking on both the attributes and then creating one Multi-forms attribute called Location and whichever attribute is required to be displayed should be selected from the below given table. This should be done whenever W_Location_Id is used.

Edit Multi-forms Attribute   ? ×

New Attribute Name: Location

W_LOCATION_D ▼

	Form Category	Display Form
Location Id	ID ▾	☐
Location Name	DESC ▾	☑ ✖

Submit   Cancel

- Also for W_SALES_CLASS_D table, Sales Class Id and Sales Class Name should be combined into Sales Class, displaying whichever attribute is required for that particular dashboard.

Edit Multi-forms Attribute   ? ×

New Attribute Name: Sales Class

W_SALES_CLASS_D ▼

	Form Category	Display Form
Sales Class Id	ID ▾	☐
Sales Class Desc	DESC ▾	☐ ✖

Submit   Cancel

- In the Customer table, Combine Customer Key and Customer Name as **Cust Id Ordered By** and display both as this will be required during creating dashboards.

Perform the above 3 whenever you see these tables being used in the dashboard creation. Also you are allowed to use your own combinations of attributes according to the data required to be displayed. These are just some examples for your understanding.

Graph Type: Maps
Geo attribute: Zip
Latitude: Zip@latitude
Longitude: Zip@longitude
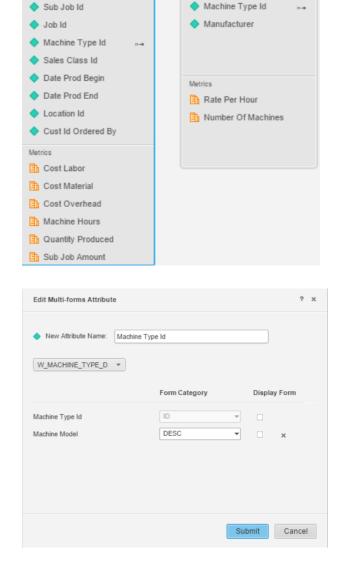**Color by**: Quantity Ordered
Size by: Number of Subjobs

Once you have the dashboard generated after using these tables, make sure you use the properties panel as well to change the map options like bubbles, pins, dynamically change zoom level. Explore all the options to get a creative dashboard.

You can now easily see in the map chart that we have derived the locations for zip codes providing information about the quantity ordered and the number of subjobs listed for that particular location. This bubble in the above visualization shows the bubble chart for the areas by the quantity ordered.

Note: Steps to create all the datasets are given in this document but the SubJob dataset is just used for mapping and for linking the attributes for a better understanding of data and is not shown in any of these visualizations.

So we will be creating a dataset with name **SubJob** as well.
Add data by exporting two tables, W_Sub_Job_d and W_Machine_Type_D as given below. Wrangle the dates and create multi-forms attribute called Machine type Id which contains Machine type Id as well as Machine model.

Remember: The visualizations that you get after performing the above steps are just for creating basic visualizations. It is expected that you explore and generate more creative/ compelling dashboards.
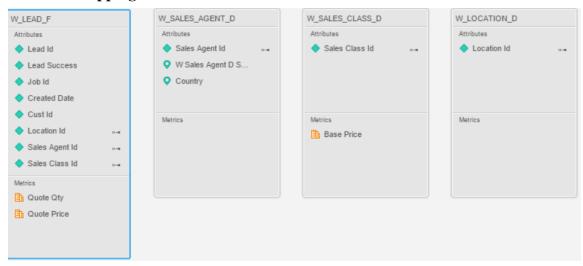
D1. Visualization 2: Lead Generation by Location and Agent

Lead Gen by Location and Agent			
Location Id	Sales Agent Id	Quote Price	Quote Qty
New York	Roman Burgess	0.74	1895500
	Connor Norton	1.57	569500
Chicago	Larissa Dickerson	1.5	1566100
	Destinee Kent	1.54	1027800
Los Angeles	Malik Hampton	1.46	2287400
	Marina Contreras	0.78	794300
	Zion Keith	0.8	181400
Seattle	Laney Maddox	0.73	2375300
	Dominick Briggs	1.57	708700
	Rory Avila	1.55	794800

Source Tables: (Save this dataset as **Lead** in the Dataset Panel)
Visualizations will be created on the basis of any dataset like Job, Lead, Sub Job, Shipment, Invoiceline. So you can always select the attributes and metrics from the dataset for creating a visualization.

W_Lead_F
W_Sales_Agent_D
W_Sales_Class_D
W_Location_D

Mapping the tables:

W_LEAD_F	W_SALES_AGENT_D	W_SALES_CLASS_D	W_LOCATION_D
Attributes	Attributes	Attributes	Attributes
◆ Lead Id	◆ Sales Agent Id ▫–•	◆ Sales Class Id ▫–•	◆ Location Id ▫–•
◆ Lead Success	◉ W Sales Agent D S…		
◆ Job Id	◉ Country		
◆ Created Date			
◆ Cust Id	Metrics	Metrics	Metrics
◆ Location Id ▫–•		🗈 Base Price	
◆ Sales Agent Id ▫–•			
◆ Sales Class Id ▫–•			
Metrics			
🗈 Quote Qty			
🗈 Quote Price			

Follow the steps as given in Visualization 1 to wrangle and combine related attributes, and always use the properties panel to display results in various formats, changing the threshold colors in properties panel also provides trends in the data displayed.

Graph Type: Grid
Rows: Location Id and Sales Agent ID
Metrics: Quote Price and Quote quantity

This information displays information about the quote price and the quantity by Sales agent for cities like New York, Chicago, Los Angeles and Seattle. The quote quantity in red denoted the lowest quote percentage for the particular city.

[**Remember:** The visualizations that you get after performing the above steps are just basic visualizations, it is expected from you to explore and generate more creative dashboards]

D1. Visualization 3: Job and Shipment Trends-To Locations

Source Tables: (Select the **Job** dataset from the dataset panel to display the **Job and Shipment Trends**)

Visualizations will be created on the basis of any dataset like Job, Lead, Sub Job, Shipment, INVOICELINE etc. So you can always select the attributes and metrics from the dataset for creating a visualization.

W_Customer_D
W_Job_F
W_Location_D
W_Sales_Class_D
W_Sales_Agent_D

Mapping the tables (In this case you have already mapped and created the Job dataset so you don't need to map it again)

Follow the Wrangle and the Multi-forms attribute instructions as given in the Visualization 1.

Graph Type: Bar Chart
Vertical: Base Price
Horizontal: Date Ship (by week) (Edit the Date Ship by attribute and create a new attribute called Date Ship by week to display the data for every week)
Color by: Location
Tool tip: Quote quantity

This bar chart displays information about the Ship dates by week for locations on basis of their base price. It is seen that the base price decreases by week in 2014.

[**Remember:** The visualizations that you get after performing the above steps are just basic visualizations, it is expected from you to explore and generate more creative dashboards]

D1. Visualization 4: Late shipment trend by number of days

Late Shipments (By # of days)	
Job Shipment Id	Ship - Days late
816593	339
815558	337
816301	337
816958	337
815261	336
815560	336
815567	336
815898	336
815571	335
812586	5
812601	5
812613	5
812614	5

Source Tables: (Save the below dataset as **Shipment** in the Dataset Panel)

Visualizations will be created on the basis of any dataset like Job, Lead, Sub Job, Shipment, INVOICELINE etc. So you can always select the attributes and metrics from the dataset for creating a visualization.

W_JOB_SHIPMENT_F
W_INVOICELINE_F
W_Customer_D
W_Location_D
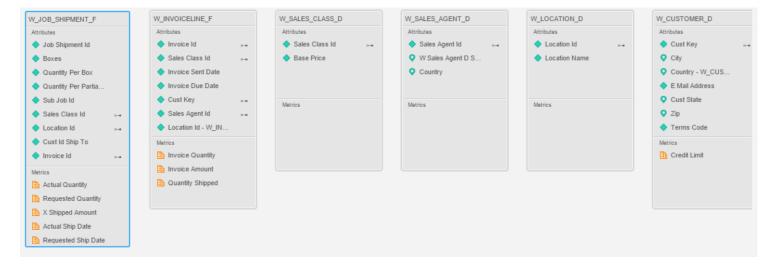W_Sales_Class_D
W_Sales_Agent_D

Mapping the tables:

Chart type: Grid
Rows: Job shipment id
Metrics: Ship days Late (Calculate the **trend for shipment days late** by considering Actual ship date and the requested ship date)

Follow the Wrangle and the Multi-forms attribute instructions as given in the Visualization 1.

This information displays the result about the days the shipment was delayed by for a particular Job Shipment id.
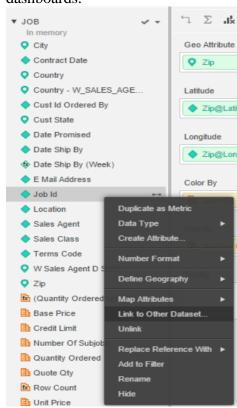
[**Remember:** The visualizations that you get after performing the above steps are just basic visualizations, it is expected from you to explore and generate more creative dashboards]
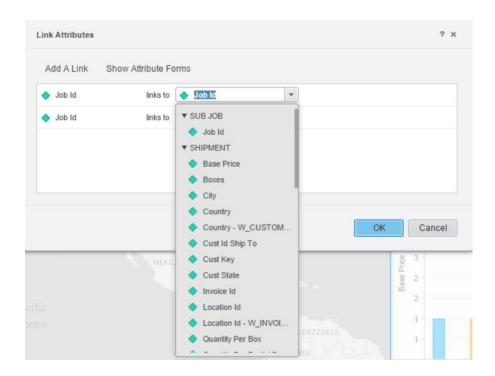
Polish (Finishing touches) Dashboard 1

- Name this Dashboard 1 as **Job and Shipment trends.**
- Provide descriptive names as titles to each visualization instead of the default titles generated
- For the map graph, use the properties to control the Bubble chart visualization.
- Apply your own preference on formatting of colors for box lines, text, and backgrounds.
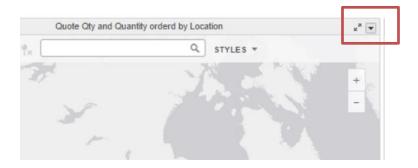- Save you work progress to the same file name used as before.

Important: Creating Interactive Dashboards (Do this for all the dashboards)
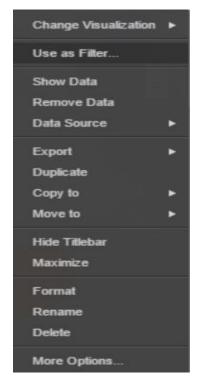
When you create the first dashboard for **Job and Shipment trends**. Always make sure you link the datasets with same attributes like Location id or Job id or any other important attribute. For example, Job and SubJob dataset have a same attribute called location id, so you can link the two datasets to see changes in the dashboards.
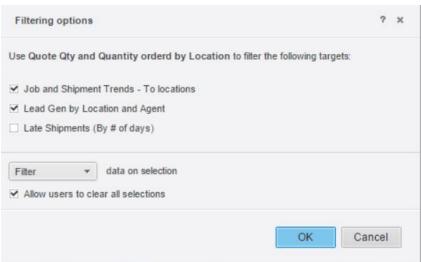
After this make sure you click on the dropdown for every visualization and click on "Use as filter…", this will display the other visualizations present in the dashboard, then check all the visualizations that you think needs to be selected for your visualization.

This will make sure that by clicking on a particular bar or bubble chart, the filter will reflect the changes in all the visualizations. Do this for all the datasets of your dashboard, so that they are linked and you get an interactive dashboard, which means that by clicking on any one graph or bubble chart, changes are reflected in other visualizations which gives you a grid report or a line and a bar chart.

Dashboard 1: Job and Shipment Trends

Answer the following questions based on Dashboard 1 for Job and Shipment Trends

- How much revenue does a company generate from its job bookings?

- How many jobs does each sales agent book?

- How many jobs have not yet shipped or have only partially shipped?

PART II. Design Dashboard 2 (D2)

Now that you have analyzed the KPIs in the one dataset and identified which locations that are performing the worst, we want to find the cause for these problems. In order to achieve this goal, we make use of the feature data blending from the additional dataset that was brought in. In this case, we will now compare the location data with the invoice trends for the Customers as well as the invoice generated for different cards.

1. Title with text boxes and images (logos) are the same as for worksheet 1, please follow steps 1 and 2 from when building out worksheet 1.

coursera	Capstone project	CU Business School
	Student: J. Doe Student No: 123-456-789	UNIVERSITY OF COLORADO DENVER

> **TIP** Duplicating a worksheet and clearing out existing visualization will simplify the effort needed to create the heading sections multiple times.

2. Create four Visualizations based on the KPIs for Dashboard 2:

D2. Visualization 1: Total invoiced amount by Sales class

Total Invoiced Amount by Class	
Sales Class Id	Invoice Amount
1 Debit Smart	$189,132,465.00
2 Credit Smart	$174,878,993.00
3 Debit NoSmart	$82,390,239.00
4 Credit NoSmart	$92,611,081.00

Source Tables: (Save this dataset as **INVOICELINE** in the Dataset Panel)
Visualizations will be created on the basis of any dataset like Job, Lead, Sub Job, Shipment, INVOICELINE etc. So you can always select the attributes and metrics from the dataset for creating a visualization.

W_INVOICELINE_F
W_Customer_D
W_Location_D
W_Sales_Class_D

W_Sales_Agent_D
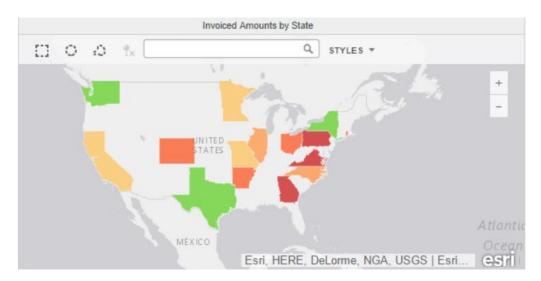
Mapping the tables:

Rows: Sales Class Id
Metrics: Invoice amount

This grid data shows information about the invoice amount for the Sales class Id for four different types of cards generated by CPI Inc.

[**Remember:** The visualizations that you get after performing the above steps are just basic visualizations, it is expected from you to explore and generate more creative dashboards]

D2. Visualization 2: Invoiced amounts by State

Source Tables: Source Tables: (Select the **INVOICELINE** dataset from the dataset panel to display the **Invoiced Amounts by State**)

Visualizations will be created on the basis of any dataset like Job, Lead, Sub Job, Shipment, INVOICELINE etc. So you can always select the attributes and metrics from the dataset for creating a visualization.

W_INVOICELINE_F
W_Customer_D
W_Location_D
W_Sales_Class_D
W_Sales_Agent_D

Chart Type: Maps
Geo Attribute: Customer State
Color By: Invoice Amount

The above map shows us the regions in United States by the invoice amounts. The regions showing green color have the highest invoice amount and the ones with red have the lowest.

[**Remember:** The visualizations that you get after performing the above steps are just basic visualizations, it is expected from you to explore and generate more creative dashboards]

TIP

Data Blending allows you to combine data from multiple, diverse sources in a single dashboard and even visualization. This is particularly useful when you are trying to analyze data that is related but maintained separately and allows you to greatly expand your analysis.

Make sure your data between the two datasets are linked on:
1. Job Id between Job and Sub Job, and
2. Sub_Job id between Sub Job and Shipment

Create any such links which you think can be joined between two tables

This worksheet also has a control built in (use as a Filter) that when you select on the grid report (top left) it update the information in the other 3 visualizations on the page

---

D2. Visualization 3: Customer Invoice Trends

Source Tables: (Select the **INVOICELINE** dataset from the dataset panel to display the **Customer Invoice Trends**)

Visualizations will be created on the basis of any dataset like Job, Lead, Sub Job, Shipment, INVOICELINE etc. So you can always select the attributes and metrics from the dataset for creating a visualization.

W_INVOICELINE_F
W_Customer_D
W_Location_D
W_Sales_Class_D
W_Sales_Agent_D

Graph Type: Bubble Chart
Vertical: Invoice Amount
Horizontal: Invoice Due Date
Color By: Customer Key

This chart displays information about the invoice amount by the invoice due dates with correspondence to the customer keys. This graph displays the information about the Customer Invoice Amount trends by Invoice due dates and the reflecting Invoice amount which helps us understand how the Invoice amount changes over the period of time for each customer.

[**Remember:** The visualizations that you get after performing the above steps are just basic visualizations, it is expected from you to explore and generate more creative dashboards]

D2. Visualization 4: Invoice Amounts by Date and Location

Invoice Amounts by Date and Location		
Location Id	Invoice Due Date	Invoice Amount
New York	1/10/2014	$5,545,428.00
Chicago	1/20/2014	$4,696,820.00
Dallas	1/8/2013	$3,597,440.00
Los Angeles	1/18/2013	$3,560,122.00
Denver	1/23/2013	$3,383,112.00
Chicago	1/21/2014	$3,373,956.00
	1/13/2013	$2,941,911.00
Dallas	1/19/2013	$2,927,780.00
Seattle	1/7/2013	$2,863,333.00
	1/12/2013	$2,815,464.00
Los Angeles	1/11/2013	$2,743,902.00
Seattle	1/12/2013	$2,735,376.00

Source Tables: (Select the **INVOICELINE** dataset from the dataset panel to display the **Invoice Amounts by Date and Location**)

Visualizations will be created on the basis of any dataset like Job, Lead, Sub Job, Shipment, INVOICELINE etc. So you can always select the attributes and metrics from the dataset for creating a visualization.

W_INVOICELINE_F
W_Customer_D
W_Location_D
W_Sales_Class_D
W_Sales_Agent_D

Rows: Location Id and Invoice Due Date
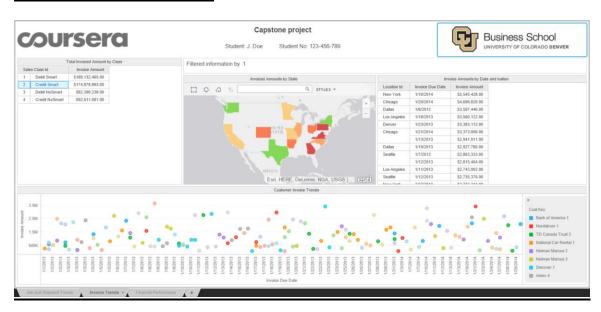**Metrics:** Invoice Amount

Similar to the previous graph, this grid data displays information about the due dates for Invoice amount by the location id.

[**Remember:** The visualizations that you get after performing the above steps are just basic visualizations, it is expected from you to explore and generate more creative dashboards]

Polish (Finishing touches) Dashboard 2

- Name this Dashboard 2 as **Invoice Trends.**
- Provide descriptive names as titles to each visualization instead of the default titles
- Apply "**Use as a Filter**" from the grid report to control the other visualizations and information on this page
- Apply your own preference on formatting of colors for box lines, text, and backgrounds.
- Save you work progress to the same file name used as before.

Dashboard 2: Invoice Trends

With Dashboard #2 completed above, select different card types and see how the rest of the information updates on the page.
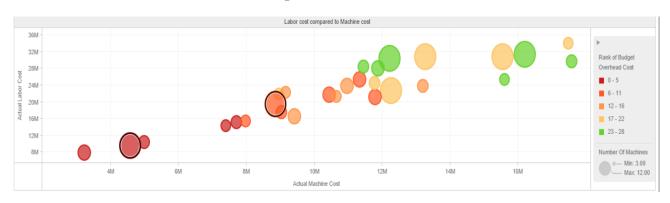
Answer the following questions based on Dashboard 2 for Invoice Trends

- Which sales class generate the highest invoice amounts?

- How many invoices are generated for a time period?

- What is the total amount invoiced for a time period?

PART III. Design Dashboard 3 (D3)

1. Create two Visualizations based on the KPIs for Dashboard 3

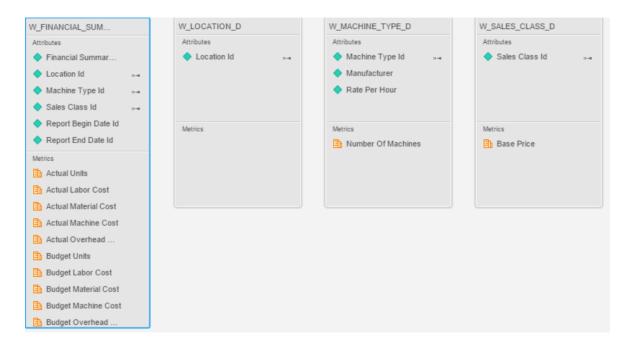### D3. Visualization 1: Labor Cost compared to Machine Cost

Source Tables: (Save the below tables as **Financial Summary Cost** dataset and display the **Labor cost compared to Machine Cost**)

Visualizations will be created on the basis of any dataset like Job, Lead, Sub Job, Shipment, INVOICELINE etc. So you can always select the attributes and metrics from the dataset for creating a visualization.

    W_FINANCIAL_SUMMARY_COST_F
    W_Location_D
    W_MACHINE_TYPE_D
    W_Sales_Agent_D

Mapping the tables:

        Chart: Bubble Chart
        Vertical: Actual Labor Cost
        Horizontal: Actual Machine Cost
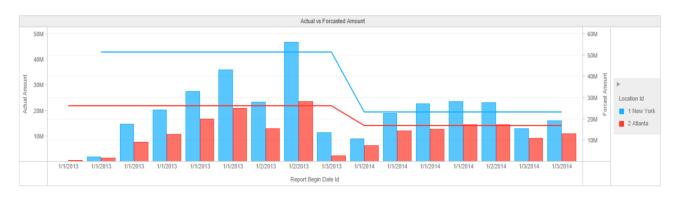   **Color by:** budget Overhead Cost
        Break by: Machine type id and Location Id
        Size by: Number of Machines

From the above bubble chart it is seen how the budget overhead cost compares in relation to the Actual machine cost and Actual labor cost. The below graph reflects the changes for the bubbles in relation with the machine type and location. The bubbles in red shows the lowest budget overhead cost and the ones in green show higher overhead budget cost.

[**Remember:** The visualizations that you get after performing the above steps are just basic visualizations, it is expected from you to explore and generate more creative dashboards]

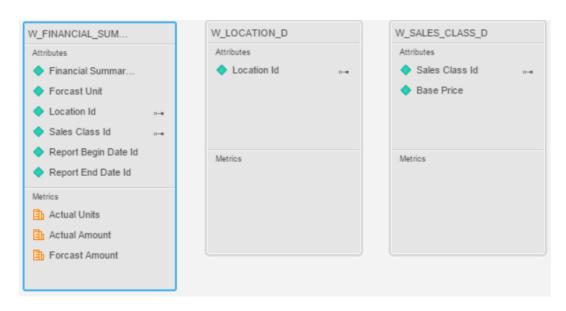**D3. Visualization 2: Actual vs. Forecasted Amount**

Source Tables: (Save the below tables as **Financial Summary Sales** dataset and display **Actual Vs Forecasted Amount**)

Visualizations will be created on the basis of any dataset like Job, Lead, Sub Job, Shipment, INVOICELINE etc. So you can always select the attributes and metrics from the dataset for creating a visualization.

> W_FINANCIAL_SUMMARY_SALES_F
> W_LOCATION_D
> W_SALES_CLASS_D

Mapping the tables:

Vertical: Actual Amount and Forecast Amount
Horizontal: Report begin Date Id
Color by: Location id

[**Remember:** The visualizations that you get after performing the above steps are just basic visualizations, it is expected from you to explore and generate more creative dashboards]

Polish (Finishing touches) Dashboard 3

1. Name this Dashboard 3 as **Financial Performance.**
2. Provide descriptive names as titles to each visualization instead of the default titles
3. Apply "**Use as a Filter**" from the Bubble Chart (top chart) to control the Actual vs Forecasted Amount combination chart (bottom chart)
4. Apply your own preference on formatting of colors for box lines, text, and backgrounds.

Dashboard 3: Financial Performance

Answer the following questions based on Dashboard 3 for Financial Performance

- Determine the location and the machine which have the highest overall machine and labor cost. Also determine which location has the lowest budget overhead cost.

- Which location is seen to have higher forecast amount in comparison to the actual amount on the basis of time period?

PART IV. Add Logos

After all the dashboards are complete, create a title bar with your Name, Coursera Logo and the UC Denver Business school logo for all the Dashboards
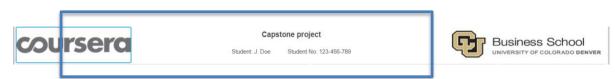
Three images (logo's) as part of the heading

The Image URL's for the images are as follow
- Top Left - Coursera: http://tinyurl.com/zxyu6we
- Top Right - CU: http://tinyurl.com/zoqmz3m

Two Text Boxes as part of the heading
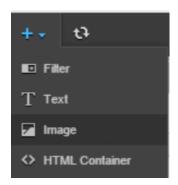

The text boxes contain the heading of the dashboard in the top and student information in the bottom.

Just click on this icon to insert text and images and edit it accordingly.
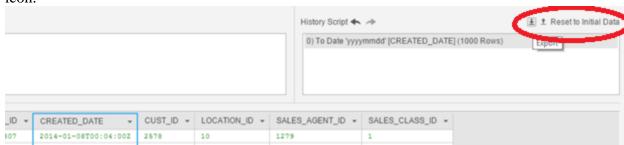

Submission of Your Work

Once your Analysis is complete the following needs to be submitted for grading

1. Submit one mstr file for the dashboards 1, 2, and 3 containing the visualizations. Use the same .mstr file and click on new sheet below to add dashboards 2, and 3 to the same .mstr file as dashboard 1.
   Name the file : **<Student Name> - Capstone project Analysis.mstr**
2. Submit your Data Wrangling Script saved during the wrangle-import process for each table and copy it in one notepad. This script can be downloaded when you wrangle the dates or names for a table. On top right you can see a download script icon.

Save the File Name: **<Student Name>_<Date>.txt**

3. Submit a document by name "<Student Name>_Project Summary.doc " containing a summary and analysis of the exercise. It should briefly cover following aspects:

   - Analysis of all the visualizations with respect to how it is helping you find the Key Performance Indicators and the causes of various problems identified.

Hint: Students can attach a screenshot of all the three dashboards in one document and can provide their analysis for each dashboard on the basis of the attributes and the metrics used.

- Conclusion of your Project and your leanings.
- Feedback: What more could you have done other than what is included in the exercise to achieve better analysis of the datasets provided to you?