

# Black Women Data 2022

## Power BI Workshop

**Case:** You are a consultant working with a human capital division that provides people analytics services to Fortune 500 organizations. The organization has asked you to build a dashboard around their team and organizational performance and employee sentiment. This case is inspired by this [Kaggle challenge](#).

**Data:** [Introduction to the HR Dataset - Version 14](#) - Last Updated April, 2021. This HR Dataset is a synthetic data set created specifically to use for HR analytics cases and is updated every year or so, by the owners. Updates include additional columns, and to make slight changes to the underlying data.

**Inspirational Questions:** Here are some open-ended questions that you can explore and try to address through creating visualizations, or R or Python analyses.

- *Is there any relationship between who a person works for and their performance score?*
- *What is the overall diversity profile of the organization?*
- *What are our best recruiting sources if we want to ensure a diverse organization?*
- *Can we predict who is going to terminate and who isn't? What level of accuracy can we achieve on this?*
- *Are there areas of the company where pay is not equitable?*
- Is there any relationship between who a person works for and their performance score?
- What is the overall diversity profile of the organization?
- What are our best recruiting sources if we want to ensure a diverse organization?
- Can we predict who is going to terminate and who isn't? What level of accuracy can we achieve on this?
- Are there areas of the company where pay is not equitable?
- Is there a relationship between age and performance
- Does working on special projects affects performance

**Data Dictionary:**

Feature	Description	DataType
Employee Name	Employee's full name	Text
EmpID	Employee ID is unique to each employee	Text
MarriedID	Is the person married (1 or 0 for yes or no)	Binary
MaritalStatusID	Marital status code that matches the text field MaritalDesc	Integer
EmpStatusID	Employment status code that matches text field EmploymentStatus	Integer
DeptID	Department ID code that matches the department the employee works in	Integer
PerfScoreID	Performance Score code that matches the employee's most recent performance score	Integer
FromDiversityJobFairID	Was the employee sourced from the Diversity job fair? 1 or 0 for yes or no	Binary
Salary	The person's yearly salary. \$ U.S. Dollars	Float
Termd	Has this employee been terminated - 1 or 0	Binary
PositionID	An integer indicating the person's position	Integer
Position	The text name/title of the position the person has	Text
State	The state that the person lives in	Text
Zip	The zip code for the employee	Text
DOB	Date of Birth for the employee	Date
Sex	Sex - M or F	Text
MaritalDesc	The marital status of the person (divorced, single, widowed, separated, etc)	Text
CitizenDesc	Label for whether the person is a Citizen or Eligible NonCitizen	Text
HispanicLatino	Yes or No field for whether the employee is Hispanic/Latino	Text
RaceDesc	Description/text of the race the person identifies with	Text
DateofHire	Date the person was hired	Date
DateofTermination	Date the person was terminated, only populated if, in fact, Termd = 1	Date
TermReason	A text reason / description for why the person was terminated	Text
EmploymentStatus	A description/category of the person's employment status. Anyone currently working full time = Active	Text
Department	Name of the department that the person works in	Text
ManagerName	The name of the person's immediate manager	Text
ManagerID	A unique identifier for each manager.	Integer
RecruitmentSource	The name of the recruitment source where the employee was recruited from	Text
PerformanceScore	Performance Score text/category (Fully Meets, Partially Meets, PIP, Exceeds)	Text
EngagementSurvey	Results from the last engagement survey, managed by our external partner	Float
EmpSatisfaction	A basic satisfaction score between 1 and 5, as reported on a recent employee satisfaction survey	Integer
SpecialProjectsCount	The number of special projects that the employee worked on during the last 6 months	Integer
LastPerformanceReviewDate	The most recent date of the person's last performance review.	Date
DaysLateLast30	The number of times that the employee was late to work during the last 30 days	Integer
Absences	The number of times the employee was absent from work.	Integer

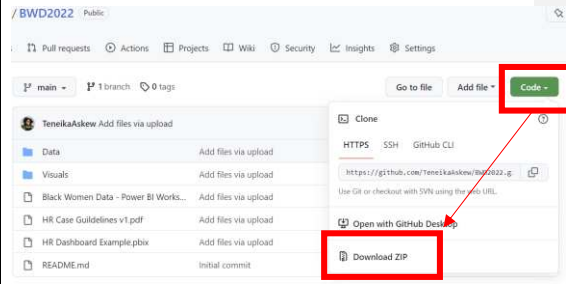
## Power BI Steps:

Outside of this workshop, you can follow the [Microsoft Power BI learning path](https://learn.microsoft.com/en-us/power-bi/fundamentals/getting-started-quickstart) to dive deeper into Power BI.

All files for this session are available on Github. You can download all files and relevant additional content there.

Once files are downloaded, you will need to unzip files to your Desktop for ease of use or a file path that is solely for projects so files paths are easy to configure.

<https://github.com/TeneikaAskew/BWD2022>



If you do not have a Microsoft account and want to use Power BI online or a virtual machine/lab without creating one,

here are training accounts I've created that you can access.  
<https://docs.google.com/spreadsheets/d/15ckDqvgIrpIvwikV-NkpA5QJ5Eb876Aa0HUK7CV-jEE/edit?usp=sharing>

### Opening Power BI Options

#### Option 1:

- Open Microsoft Virtual Machine
- Visit Become a Power BI Analyst Path
- Open up this [lab](https://learn.microsoft.com/en-us/training/modules/get-data/lab-prepare):
  - <https://learn.microsoft.com/en-us/training/modules/get-data/lab-prepare>
- Sign in using your own account or a training account
- Click on Start Lab
- Lab environment will begin to build and open to a Windows environment where Power BI is readily available

### Lab - Prepare data in Power BI Desktop

45 minutes


This unit includes a lab to complete.

Use the free resources provided in the lab to complete the exercises in this unit. You will not be charged.

Microsoft provides this lab experience and related content for educational purposes. All presented information owned by Microsoft and intended solely for learning about the covered products and services in this Microsoft module.

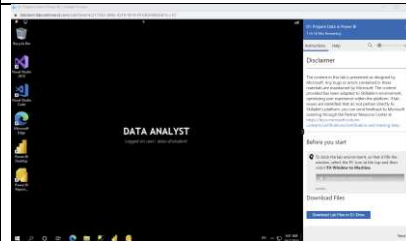
[Sign in to launch the lab](#)

### Access your environment

 Your lab environment is being built

Your lab will be ready in about 2 minutes.

Lab Environment will look like this. The environment expires in 4 hours. If you would like to save your work, I suggest you log into a browser and sign into your Google drive and save your Power BI package there.



Option 2:

- Download Power BI Desktop
- <https://www.microsoft.com/en-us/download/details.aspx?id=58494>
- You can download Power BI through the download option that redirects to the **Microsoft App Store** or download the executable in **Advanced download options**.



**Microsoft Power BI Desktop**

With the Power BI Desktop you can visually explore your data through a free-form drag-and-drop canvas, a broad range of modern data visualizations, and an easy-to-use report authoring experience.

Download >

Advanced download options >

Option 3:

- Power BI online is available, but this version will look different from the Desktop version. <https://app.powerbi.com/>



**Getting Started**

Open Power BI and Get Data

- To simplify this process we will use one data source: HR Dataset v14
- Select Text/CSV
- In the File Dialog, navigate to **HRDataset\_v14**
  - You should only use the Excel Workbook option should you fall behind and need the tables we will build in the lab

**Power BI Desktop**



Get data

Recent sources

Name

HRData\_Employees

HRDataset\_v14

Preview the file then click Transform Data to begin data preparation

#### HRDataset\_v14.csv

File Origin: (65001: Unicode (UTF-8)) Delimiter: Comma Data Type Detection: Based on first 200 rows

Employee_Name	EmpID	MarriedID	MaritalStatusID	GenderID	EmpStatusID	DeptID	PerfScoreID	FromDiversityJob
Adinolfi, Wilson K	10026	0	0	1	1	5	4	
Alt, Sid, Karthikeyan	10084	1	1	1	5	3	3	
Akinkuole, Sarah	10196	1	1	0	5	5	3	
Alagbe, Trina	10088	1	1	0	1	5	3	
Anderson, Carol	10089	0	2	0	5	5	3	
Anderson, Linda	10002	0	0	1	1	5	4	
Andreola, Colby	10194	0	0	0	1	4	3	
Athwal, Sam	10062	0	4	1	1	5	3	
Bachiochi, Linda	10114	0	0	0	3	5	2	
Baccong, Alejandro	10250	0	2	1	1	3	3	
Baccenki, Rachael	10252	1	1	0	5	5	3	
Barbosa, Thomas	10242	1	1	1	5	5	3	
Barbosa, Hector	10012	0	2	1	1	3	4	
Barone, Francesco A	10065	0	0	1	1	5	3	
Barton, Nader	10066	0	2	1	3	3	3	
Bates, Norman	10061	0	0	1	4	3	3	
Beak, Kimberly	10023	1	1	0	2	5	4	
Beatrice, Courtney	10055	0	0	0	1	5	3	
Becker, Renee	10245	0	0	0	4	3	3	
Becker, Scott	10277	0	0	1	3	5	3	

Extract Table Using Examples Load Transform Data

Begin with duplicating the table. Right click on the table and click **Duplicate Table**

Queries [1]

HRDataset\_v14

- Copy
- Paste
- Delete
- Rename
- Enable load
- Include in report refresh
- Duplicate
- Reference
- Move To Group
- Move Up
- Move Down
- Create Function...
- Convert To Parameter
- Advanced Editor
- Properties...

Right Click on the table and name this Manager

Remove other columns except ManagerName, Manager ID and EmpID

Go to the original table and rename this table from HR Dataset v14 to Employees	
Remove ManagerName from Employees	
Repeat steps to duplicate the Employees table.	
Delete Position from original data table	



Add Fully Loaded Cost

File Home Transform Add Column View

Column From Custom Invoke Custom  
Examples Column Function

Conditional Column  
Index Column  
Duplicate Column

General

### Custom Column

Add a column that is computed from the other columns.

New column name  
Fully Loaded Costs

Custom column formula

= [Salary]\*1.25

Available columns  
Salary  
PositionID  
State  
Zip  
DOB  
MaritalDesc  
CitizenDesc

[Learn about Power Query formulas](#)

✓ No syntax errors have been detected.

**Commented [TA1]:** When you think about adding a new employee to your payroll, determine what the actual financial cost of doing so means to your business. This includes the dollars and cents over and above the basic wage or salary you agree to pay. There's a rule of thumb that the cost is typically 1.25 to 1.4 times the salary, depending on certain variables. So, if you pay someone a salary of \$35,000, your actual costs likely will range from \$43,750 to \$49,000. Some added employment costs are mandatory, while others are a little harder to pin down. Fortunately, there may be tax savings to offset some of the costs.

Update Zip Code column to ensure leading zero is back. We will use the Column from Examples feature.

Name this column Zip\_c

**Add Column From Examples**

Enter sample values to create a new column (Ctrl+Enter to apply).

Transform: `Text.PadStart(Text.From([Zip], "en-US"), 5, "0")`

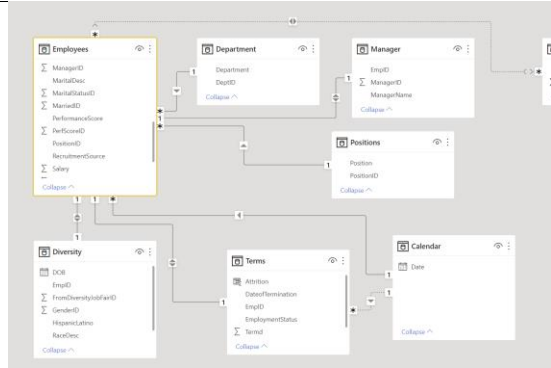
	PositionID	State	Zip	DOB	MaritalDesc	Custom	
1	29	MA		1960	7/10/1983	Single	01960
2	27	MA		2148	5/5/1975	Married	02148
3	20	MA		1830	8/18/1888	Married	01830
4	29	MA		1886	8/22/1888	Married	01886
5	29	MA		2169	8/8/1889	Divorced	02169
6	29	MA		1844	5/22/1977	Single	01844
7	24	MA		2230	5/24/1979	Single	02130
8	29	MA		2299	2/18/1963	Widowed	02299
9	29	MA		1902	2/11/1970	Single	01902
10	24	MA		1886	1/27/1888	Divorced	01886
11	29	MA		1902	1/12/1974	Married	01902
12	29	MA		2062	2/21/1974	Married	02062
13	9	TX		78230	7/4/1888	Divorced	78230
14	29	MA		1830	7/20/1883	Single	01830
15	29	MA		2747	7/15/1977	Divorced	02747
16	29	MA		2050	10/18/1981	Single	02050

### Modeling Data

In this exercise we will not need to model the data as Power BI has autodetected the relationships.

In future dashboards you may need to create the relationships between tables.





Now we are done in Power Query and move to the model. The first step is to update Salary and fully loaded cost -change type to currency.

**Power BI Desktop - HR Dashboard Example**

**Table tools | Column tools**

Name: Salary  
Data type: Whole number  
Format: Whole number  
Summarization: Sum  
Data category: Uncategorized

Employee_Name	EmpID	MarriedID	PerformanceScore	PerfScoreID	PositionID	Salary
Alagbe, Trina	10088	1	5	3	3	64,981
Anderson, Carol	10069	0	5	5	3	50,825
Athwal, Sam	10062	0	2	5	3	58,385
Bachiochi, Linda	10114	0	2	5	3	47,837
Baczanski, Rachael	10252	1	1	5	3	54,670
Barbara, Thomas	10242	1	1	5	3	47,211
Barone, Francesco A	10265	0	0	2	5	58,709
Barton, Nader	10066	0	2	5	3	52,505
Bates, Norman	10061	0	0	4	5	57,834
Becker, Scott	10277	0	0	3	5	53,250
Bernstein, Sean	10046	0	0	2	5	51,044
Biden, Lowan M.	10226	0	2	2	5	64,919
Brill, Donna	10177	1	1	5	3	53,492
Bugali, Josephine	10303	0	3	3	5	64,375
Carry, Michael	10115	0	0	2	5	52,646

**Table tools | Column tools**

Name: Fully Loaded Costs  
Data type: Text  
Format: Text  
Summarization: Sum  
Data category: Uncategorized

Formatting options: Whole number, Decimal number, Fixed decimal number, Date/time, Date, Time, Text, True/false, Binary.

0

Wednesday, March 4, 2015

0

### Data type change

With this data type change, your data will be stored differently. This may cause a loss of data or precision. After you make this change, you can restore the column by refreshing the table.

Do you want to continue?

OK Cancel

0

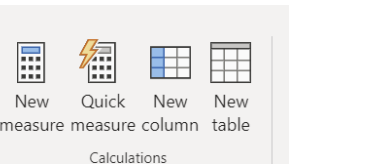
Thursday, February 18, 2021

0

### Add a Tenure field to the model using DAX

Tenure =

`DATEDIFF('Employees'[DateofHire], TODAY(), YEAR)`



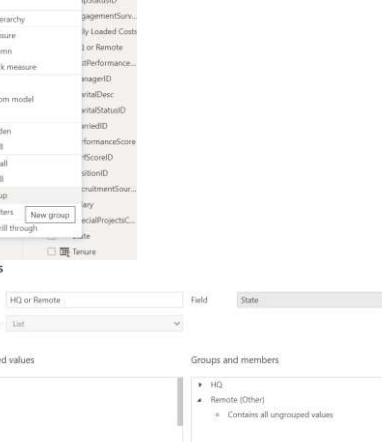
Calculations

1 Tenure = DATEDIFF('Employees'[DateofHire], TODAY(), YEAR)

RecruitmentSource PerformanceScore EngagementSurvey

### Add a HQ or Remote field to the model using Groups

- Right-click on State
- Click on **New Group**
- Click on MA then Group
- Rename the Group **HQ**
- Check **Include Other group**
- Rename the Group to **Remote**



Groups

Name: HQ or Remote Field: State

Group type: List

Un grouped values

- AL
- AZ
- CA
- CO
- CT
- FL
- GA
- ID
- IL
- IN
- KY

Groups and members

- HQ
- Remote (Other)
  - Contains all ungrouped values

Group Un group Include Other group (C)

OK Cancel

Add Attrition Field to **Terms** table  
Attrition = `CALCULATE(COUNTA(Terms[Termd]), Terms[Termd]=1)/COUNTA(Terms[Termd])`

Structure		Formatting		Properties	
✓ 1 Attrition = CALCULATE(COUNTA(Terms[Termd]), Terms[Termd]=1)/COUNTA(Terms[Termd])					
Termd	DateofTermination	TermReason	EmploymentStatus	Attrition	
0172	0	N/A-StillEmployed	Active		
0127	0	N/A-StillEmployed	Active		
0174	0	N/A-StillEmployed	Active		
0135	0	N/A-StillEmployed	Active		
0010	0	N/A-StillEmployed	Active		
0043	0	N/A-StillEmployed	Active		
0271	0	N/A-StillEmployed	Active		
0084	1 Thursday, June 16, 2016	career change	Voluntarily Terminated	0.3%	
0196	1 Monday, September 24, hours		Voluntarily Terminated	0.3%	
0069	1 Tuesday, September 6, 2	return to school	Voluntarily Terminated	0.3%	
0252	1 Thursday, January 12, 20	Another position	Voluntarily Terminated	0.3%	
0242	1 Monday, September 19,	unhappy	Voluntarily Terminated	0.3%	
0066	1 Thursday, April 6, 2017	Another position	Voluntarily Terminated	0.3%	
0267	1 Friday, April 4, 2014	career change	Voluntarily Terminated	0.3%	

Add a calendar table in the model using **DAX**  
Calendar = `CALENDAR(FIRSTDATE('Employees'[DateofHire]), LASTDATE('Employees'[DateofHire]))`

Structure		Calendars	Relationships	Calculations
1 Calendar = CALENDAR(FIRSTDATE('Employees'[DateofHire]), LASTDATE('Employees'[DateofHire]))				
Data				
1/6/2006 12:00:00 AM				
1/20/2006 12:00:00 AM				
1/11/2006 12:00:00 AM				
1/12/2006 12:00:00 AM				
1/13/2006 12:00:00 AM				
1/14/2006 12:00:00 AM				

## Visualizing Data

- Select your color palette for the report
- In this example we use Classic
  - You can also upload custom themes
- Add a grey background
- Select the page
  - Click the format icon
  - Click **Wallpaper**
  - Select Grey

This report

Power BI

Browse for themes  
Theme gallery  
Customize current theme  
Save current theme

Visualizations

Format page

Search

Page information

Canvas settings

Canvas background

Wallpaper

Color

Image

Browse...

Image fit

Normal

Transparency

0 %

Reset to default

Fields

Search

Calendar

Department

Diversity

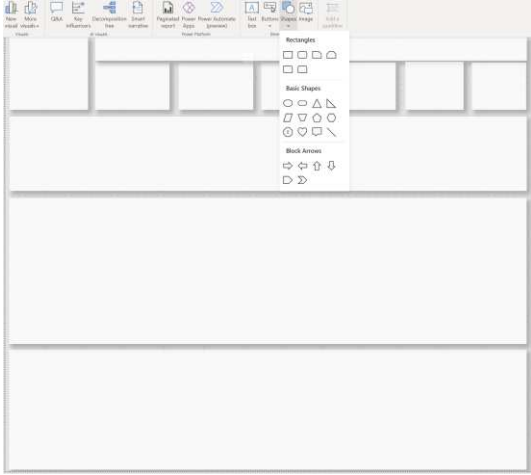

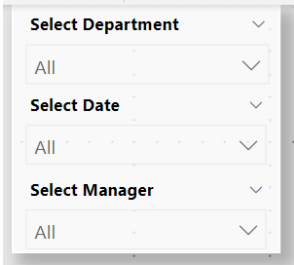


Employees

Manager

Positions

Terms

Create your layout using the shapes option and begin to highlight your page layout of

	
<p>Add a Title</p> <ul style="list-style-type: none"> <li>Select the text box and add it to the top shape</li> </ul>	
<p>Slicers (Filters)</p> <ul style="list-style-type: none"> <li>Select the slicer visual</li> <li>Check Department</li> <li>Repeat for Date and Manager</li> </ul>	
<p>Hero Stats / BANS</p> <ul style="list-style-type: none"> <li>Select the Card visual</li> <li>Check Fully Loaded Costs</li> <li>Repeat for Employee ID, Absences, Tenure (Median), Avg Performance Score</li> </ul>	
<p>Location of Employees and Recruitment Sources</p> <ul style="list-style-type: none"> <li>Check Fully Loaded Costs</li> <li>Repeat and check Recruitment Source</li> </ul>	
<p>Employee Satisfaction</p> <ul style="list-style-type: none"> <li>Drag Department to Y-axis</li> <li>Drag Emp ID to X-axis</li> </ul>	

- Add Performance Score to Legend
- Rename EmpSatisfaction to Rating in visual
- Exclude executive office from satisfaction. May want to exclude executive office with only 1 person in that office



#### Performance by Salary and Tenure

- Add Salary to X-axis
- Add Tenure to Y-axis
- Add Performance Score to Legend



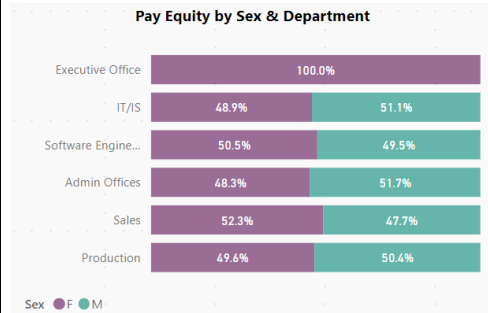
#### Direct Report Performance by Manager

- Add Manager Name to X-axis
- Add Performance Score to Y-axis
- Add Performance Score to Legend



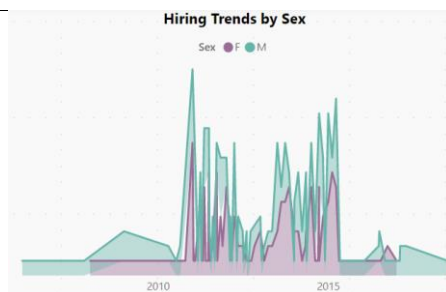
#### Pay Equity by Sex and Department

- Add Department to X-axis
- Add Salary to Y-axis
- Add Sex to Legend



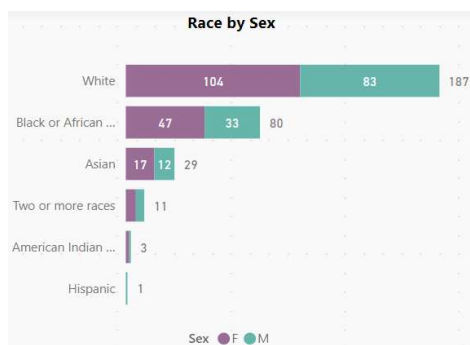
#### Hiring Trends by Sex

- Add Date to Hire to X-axis
- Add Employees to Y-axis
- Add Sex to Legend

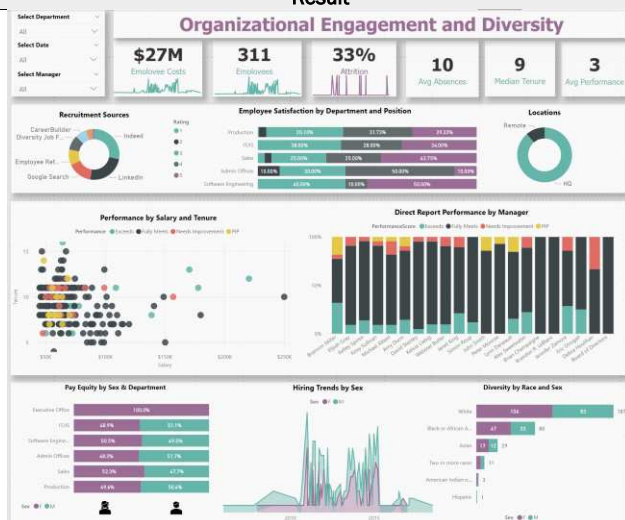


## Diversity by Race and Sex

- Add Race Desc to X-axis
- Add Employees to Y-axis
- Add Sex to Legend



Result
<p>                     The results of the study show that the proposed method is effective in detecting and classifying anomalies in network traffic. The model achieved a high accuracy of 95% in identifying malicious traffic, while maintaining a low false positive rate of 2%. The analysis of the results indicates that the model is particularly effective in detecting zero-day attacks and malware traffic, which are often challenging to detect using traditional signature-based methods. The model's performance was consistent across different network environments and traffic volumes, demonstrating its robustness and scalability. The results also show that the model is able to adapt to new types of anomalies, suggesting its potential for long-term use in network security. The study concludes that the proposed method is a promising approach for enhancing network security and detecting malicious traffic.                 </p>





Inspiration:

<https://www.kaggle.com/code/abdelrahmanralarqan/hr-visualization/notebook>

[https://public.tableau.com/app/profile/ed.myers/viz/ExploratoryDashboard\\_16183795969740/Dashboard2](https://public.tableau.com/app/profile/ed.myers/viz/ExploratoryDashboard_16183795969740/Dashboard2)

<https://public.tableau.com/app/profile/decisive.data/viz/HumanResources-HR9Box/9BoxEmployeeRatings>

[https://public.tableau.com/app/profile/techknomatic.services.pvt.ltd./viz/AttritionDashboard\\_15576673167290/AttritionAnalysis](https://public.tableau.com/app/profile/techknomatic.services.pvt.ltd./viz/AttritionDashboard_15576673167290/AttritionAnalysis)

<https://public.tableau.com/app/profile/ajay.varghese/viz/RWFD1-HRDASHBOARD/Overview>

<https://public.tableau.com/app/profile/spencer.baucke/viz/The500HRClub/The500HRClub>

