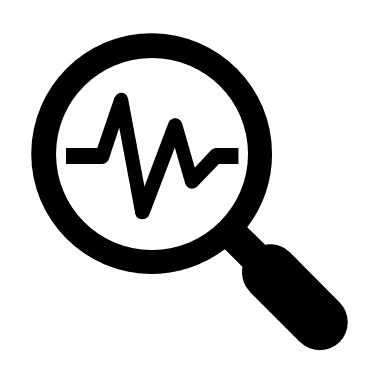
DSA 5900-001

Final Report for Professional Practice, Fall 2020



**Data Analysis and Visualization**

of

A picture containing drawing

Description automatically generated

**Course Evaluation System**

&

Text

Description automatically generated

**Employees Attendance Portal**

by

**Krishna Ganne**

([krishnaganne@ou.edu](mailto:krishnaganne@ou.edu))

OU NET ID: 113477333

Credit Hours: 4

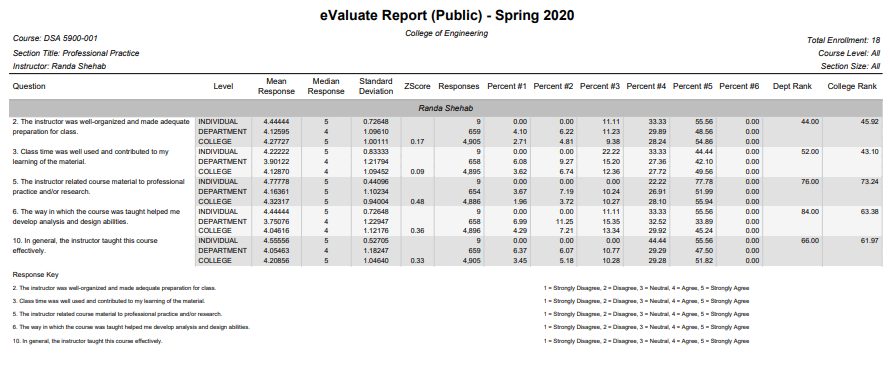
**Supervisor:** Kevin Buck, Associate Director of OU IT Learning Spaces, University of Oklahoma

**Mentor:** Aaron Biggs, Executive Director of Technology Advancement, Office of the Senior Vice President and Provost, University of Oklahoma

**Course Evaluation System**

**Introduction:**

The **eValuate** portal of University of Oklahoma is a web-based course evaluation system started in Spring 2009 to replace the process of converting the paper evaluation forms into evaluation reports. This new system has not only decreased cost and waste drastically but also gave faster feedback of the course evaluations to faculty. Moreover, the online system eliminated the need for utilizing precious class time to complete paper evaluations and eliminated the occasional loss or mistakes that occurred previously. Following the guidelines of the policy adopted by The Faculty Senate and the President since 1994, a mandatory tabular report that summarizes the student’s responses to five key questions for each course is submitted to UOSA and they are publicly available at <https://ou.edu/provost/course-evaluation-data>

  
*Figure 1: The existing course evaluation report available for public*

1. **Business Understanding:**

Although the existing reports serve the purpose of summarizing the student’s response for the course but give little if not no insights for the end user especially for instructors. Moreover, there has not been any analysis on the feedback submitted by the students, and it becomes a painful task to go through each one of them to know if the overall feedback of the course is good and to check if it is improved over the years. The main objective of the project is to develop an app to see student’s feedback for the course in a visually appealing way, interactive and giving insights about the course.

When the due date for the course feedback submission is reached, students response is consolidated and run on eValuate reporting engine. These reports are made available on eval.ou.edu website for the instructor’s viewing and an email is sent to respective instructors. Out of these reports, exactly 5 questions and their responses are available on OU’s Provost website for public viewing conforming to the federal policy. The dataset for the project is pulled together from these public reports and the resulting visualization app is secured through Microsoft 365 Business Premium subscription of university data adhering to data compliance. Although the end user of the app is public but due to “Publish to web” restrictions on the app, it is used only by instructors and students at the University of Oklahoma. Therefore, the visualization of the data is being made catering to the needs of students, and instructors.

After comprehensive meetings with students, and instructor, the following are the important requirements to be included:

1. Visualization must be simple and easy to understand.
2. Instructors should be able to choose what they want to see. In other words, highly interactive visualization.
3. Data Analysis that might be complex under the hood but easy to interpret when presented.
4. Avoid giving too much information i.e., only important measures.
5. Ability to reset the selection and look for some other entry.
6. To way to look at how they have performed over the years.
7. Tabulated data giving a chance to look at the actual feedback.

1. **Data Understanding:**

The dataset for the project is publicly available on OU’s provost website for the course as old as 2006 but due to evaluation questionnaire changes, data inconsistency, and with little or no major significance of legacy data, I took historical data of last 3 academic years i.e., 2017-18, 2018-19, and 2019-20 for all the semesters including Summer but the app is nowhere confined to only these years as it’s scalable.

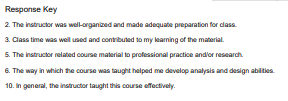
The initial approach of the project was to embed the visualization in the evals.ou.edu website but due to confidentiality, corrupting the existing system in place, and no approval from Provost, we had to approach a different solution. The entire university including faculty have access to Office 365 under Microsoft Licence so it was decided to accomplish the task using Power BI as it’s free, secured and easily accessible. The entire project is done on Power BI Desktop making use of Power Query for data transformation, Data Analysis Expressions (DAX) functional language for data analysis, BI Service for deployment and publishing.

The following tables illustrates the data columns and their description:

**Course Data:**

|  |  |
| --- | --- |
| **Title** | The title of the report lists the college or program the report belongs to |
| **Zap Number** | A unique number generated by OU IT which identifies department, section, and instructor for each course |
| **Semester** | The semester for which evaluations were filled out will appear in the top right corner of the report |
| **Course** | Department code, course number, and section number |
| **Title** | Title of the course |
| **Course Level** | The course levels considered for the report. This may be one level, multiple levels, or all and is determined by the college. |
| **Instructor** | Name of instructor who taught the evaluated section |
| **Class Size** | The class sizes considered for the report. This may be one size, multiple sizes, or all. |
| **Size** | is arbitrary and is defined by each college individually. |
| **Enrolment Count** | Number of students enrolled in the evaluated section |

**Questionnaire Data:** The questionnaire for the course mostly depends on the department it belongs to so not every course has the same questions yet those are to be included in the report for better understanding of the measures.



**5-Point Likert Scale** for the above questions:



**Responses to questions are reported in the following ways:**

|  |  |
| --- | --- |
| **INDIV** | Responses for only the section listed at the top of the report |
| **DEPT** | Responses within the department listed at the top of the report |
| **SIML** | Responses from sections with a similar course level and size within the college |
| **COLL** | Responses within the college |

**The following information is given regarding the responses to each question:**

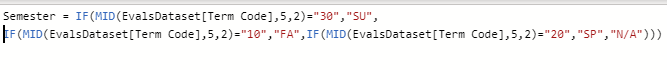
|  |  |
| --- | --- |
| **Mean Response** | The average student response to a question |
| **Median Response** | The value of the middle response to each question, separating the higher half of responses from the lower half |
| **Standard Deviation** | The average response’s deviation from the mean response. A smaller standard deviation signifies responses were clustered nearer to the mean, while a larger standard deviation signifies that responses were more distant from the mean response. |
| **Number of Responses** | The number of recorded responses for this question by each level |
| **Z‐Score** | The z‐score expresses a score in terms of the number of standard deviation units the raw score is below or above the mean of the distribution. |
| **Similar z‐score is computed by** | (similar mean‐course mean) DIVIDED BY similar standard deviation College z‐score is computed by, (college mean‐course mean) DIVIDED BY college standard deviation Note, The z‐score does not appear on all reports |
| **Percent** | These columns list how respondents answered, showing the percent of responses to each answer marked for a given question. |
| **Department Rank** | The percentile in which the instructor’s responses fall, within the department. Example, A score of 57 signifies that the instructor scored better than 57% of instructors within the department. |
| **College Rank** | The percentile in which the instructor’s responses fall, within the college.   Example,  A score of 85 signifies that the instructor scored better than 85% of instructors within the college. |

1. **Data Preparation and Modelling:**

Although there are many fields available for data analysis, only the most significant and valuable ones (based on how effective it conveys the message to the end user) are selected at the end. Moreover, the measures that are used in the overall visualization should not misled or overwhelm the end user as not all are aware of the statistics jargon, so these following measures for the interactive visualization that gives just what the user needs and the limitation of the visual space on a single page:

*Mean, median, standard deviation, responses, and other KPIs across Instructor, Department and College.*

Additionally, custom columns are created to better make sense of the attributes something like below:

  
*Figure 2: A sample of DAX code creating an additional column for better visualization and understandability.*

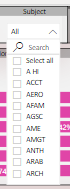
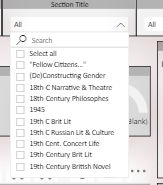
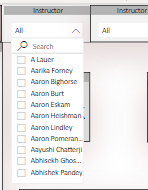
Once the data pre-processing step is done, the data is loaded to the model following a set of instructions in Power Query, the M functional language of Microsoft as follows:



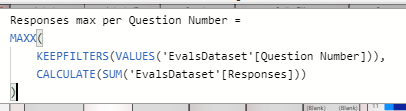
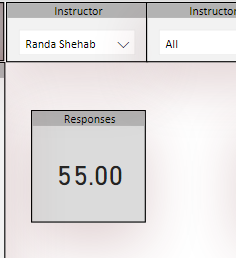
1. **Data Visualization:**

There has been multiple visualization and analysis of the eVals data before but the one that implements the requirements is at most effective, visually appealing, and catering to the needs of the instructor hence the following visuals are selected.

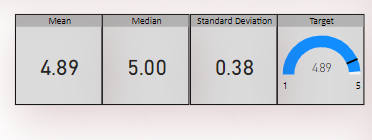
*Requirement#2 :Instructors should be able to choose what they want to see. In other words, highly interactive visualization.*



*Requirement#3: Data Analysis that might be complex under the hood but easy to interpret when presented.*

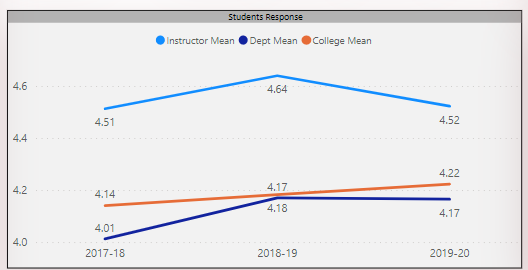
*Requirement#4: Avoid giving too much information i.e., only important measures.*



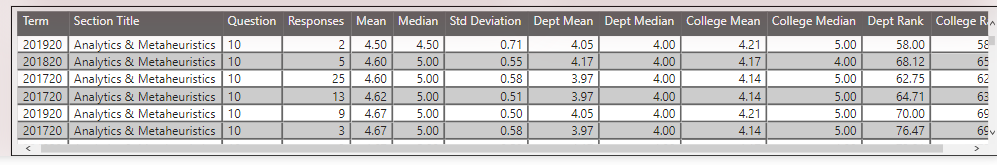
*Requirement#5: Ability to reset the selection and look for some other entry.*



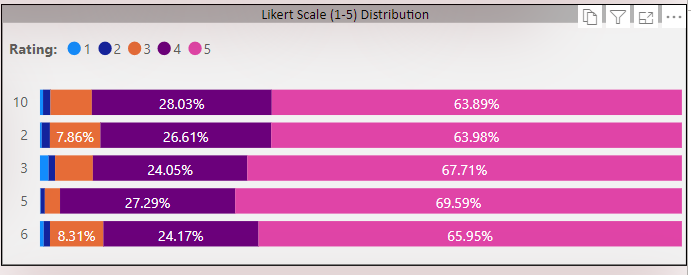
*Requirement#6: To way to look at how they have performed over the years.*



*Requirement#7: Tabulated data giving a chance to look at the actual feedback.*



*Requirement#1: Visualization must be simple and easy to understand.*



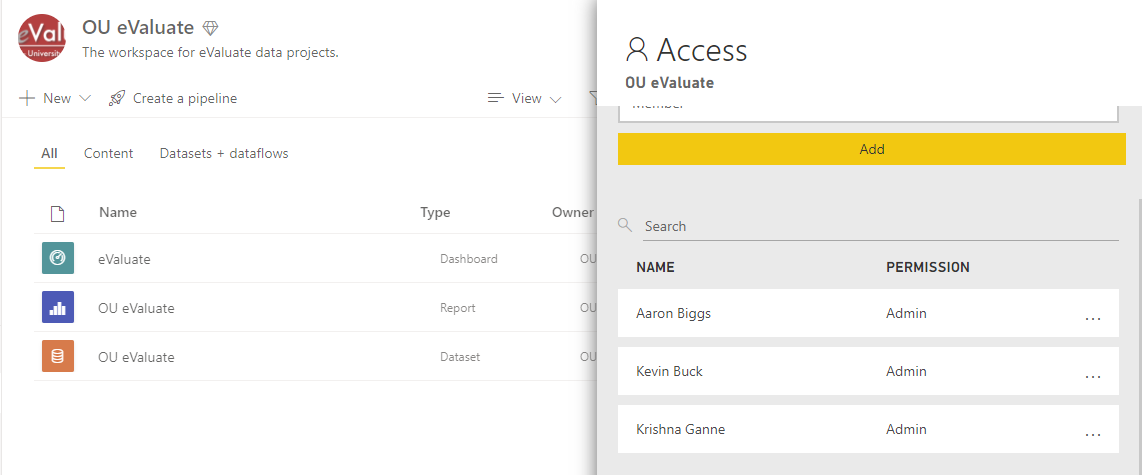
1. **Data Evaluation:**

Once the visualizations are ready, a test plan is created to test all the columns and measures by using automated DAX and otherwise, manually. The following tests are done to ensure the data accuracy.

|  |  |  |
| --- | --- | --- |
| **Test Scenario** | **Test Case** | **Test Status** |
| Data Loading | Verify the data source pointed to Learning Spaces sharepoint folder. | **PASS** |
|  | Check if the Sharepoint and Power BI are connected well and there are no access issues. | **PASS** |
|  | Verify that all columns are loaded. | **PASS** |
|  | Verify that the record count is matched between Sharepoint file and Power BI Power Query. | **PASS** |
|  | Check for non-fatal errors. | **PASS** |
| Visualization | Display date and time. | **PASS** |
|  | Display ‘0’ instead of ‘(Blanks)’ for measures like Mean, Median, Standard Deviation | **PASS** |
|  | Compare the Likert Scale distribution and the source data. | **PASS** |
|  | Check if the selection filters all the visuals | **PASS** |
|  | Compare the line chart visuals and the tabular visual. | **PASS** |
| Deployment | Check the report access permissions. | **PASS** |
|  | Verify if all the visuals are published to Service. | **PASS** |
|  | Check if the app is accessible only to the organization. | **PASS** |
|  | Check if the “Analyse to Excel” is disabled. | **PASS** |

1. **Deployment and publishing:**

Once the visualizations were SIT tested, it was sent to the end users for Beta Testing. There were no issues reported while testing it, so the above data project has been deployed over the cloud using Microsoft Power BI Service. A new workspace is created with the name “OU eValuate” adding Aaron Biggs as an Administrator for data refresh in the future.

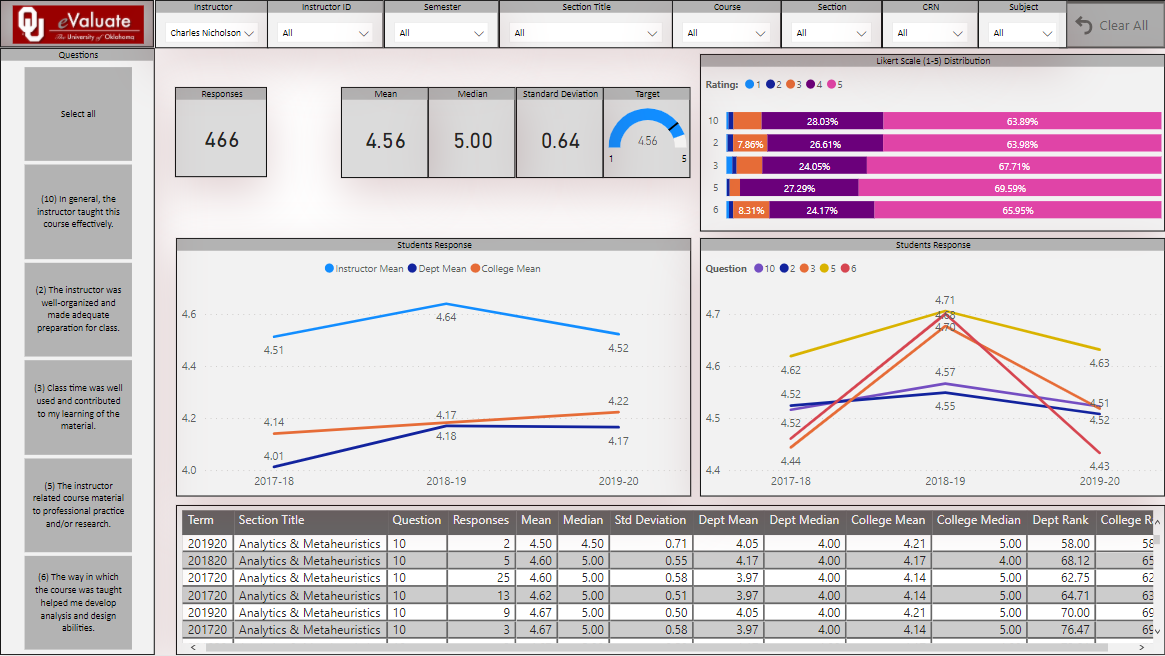


The access permissions and ownership is set and it’s published for all OU users accessible on Provost website and as an App.



1. **Results:**

The end user accesses the visualization through the website link or app, and the below visually appealing information pops up. Now the user can search the feedback of the course by choosing any of the entries: Instructor, Instructor ID, semester, course name (section title), course number, section, subject, or even CRN as below.



Based on their selection, the appropriate questions are filtered and displayed. In addition, the student’s response for each question is presented by various charts like bar chart to reflect the distribution of ratings, line chart to give insights to the instructor how they have performed over the years as well as in comparison with the department and college.

1. **References:**

* OU Course Evaluation Guidelines

( <https://ou.edu/provost/faculty-affairs/evaluation>)

* Reza Rad,2019: *Power BI from Rookie to Rock Star, Book 4: Power BI Modeling and DAX*
* Microsoft Power BI Desktop Documentation

(<https://docs.microsoft.com/en-us/power-bi/fundamentals/desktop-getting-started>)

* Microsoft Power BI Service

([*https://docs.microsoft.com/en-us/power-bi/admin/service-admin-rls*](https://docs.microsoft.com/en-us/power-bi/admin/service-admin-rls)*)*

* Reza Rad Blog for Row-level Security:

*(*[*https://radacad.com/dynamic-row-level-security-in-power-bi-with-exclude-and-include-rules*](https://radacad.com/dynamic-row-level-security-in-power-bi-with-exclude-and-include-rules)*)*

* Microsoft Data Preparation Strategies

([*https://powerbi.microsoft.com/en-us/blog/tag/data-preparation/*](https://powerbi.microsoft.com/en-us/blog/tag/data-preparation/)*)*

**OU IT Employees Attendance Portal**

**Introduction:**

The Information Technology of the University of Oklahoma (OU IT) form the backbone of the school by creating and maintaining the best possible learning environments through world-class IT services and Learning Spaces and Services, a part of OU IT, have been pioneering in the mission by supporting academics with updated, standardized, and user-friendly technology in around 150 centrally scheduled classrooms. Both departments having close to 100 student employees (PTEs) mentored by highly talented and innovative full-time employees make this possible. But there was no employee’s portal to track PTE’s timesheets, attendance infractions, time-off requests, or a way for the full-time employees to appraise their performance over the semester and subsequent pay raise at the beginning of the immediate semester. Moreover, it was the frustration of students to keep track of their timesheet correction all the time hence the necessity to have a portal.

1. **Business Understanding:**

Student employees access their timesheet on WhenIWork software as well as WorkForce software, their personal information through Qualtrics survey. There is no single platform to have everything they need for effective functioning working with OU IT. The foremost objective of the project is to give the employees access to their logged data securely without compromising their right to privacy. At the same time, giving insights to make them well-informed about their shortcomings and thereby committing no infractions while working for OU IT. Although there are n ways to do it, my objective is to create the portal with the resources available without requiring additional licences or costs hence I chose Power BI with other Office 365 applications which are free, secured, and readily available for all employees and students of OU. For ETL( Extract-Transform-Load) of multiple sources like WhenIWork software, Qualtrics, SharePoint and other internal applications, I used Power Query, the M functional language and for data analysis, I used Data Analysis Expression (DAX) functional language, and for deployment, publishing, and maintenance, I made use of Power BI Service.

1. **Data Understanding:**

The dataset for the project is pulled primarily from WhenIWork and Qualtrics software and placed in the Microsoft SharePoint site which acts as the only source for all the subsequent steps. The following tables shows the data columns and their description:

**Employees Personal Information Table:**

|  |  |  |
| --- | --- | --- |
| **First Name** | **Last Name** | **OU Email Address** |
| **OU 4x4** | **Sooner ID** | **EMPLID** |
| **Planned Graduation** | **Classification** | **Position** |
| **DOB** | **Phone Number** | **Major** |

**Employees WhenIWork Schedules Table:**

|  |  |
| --- | --- |
| **Date** | Shift Date |
| **Employee** | Employee Full Name |
| **Employee ID** | EMPLID |
| **Position** | Position at Learning Spaces |
| **Schedule** | Schedule name |
| **Site** | Location |
| **Start** | Shift start time |
| **End** | Shift end time |
| **Unpaid Break** | Allowed break time duration |
| **Total Hours** | Total hours scheduled |
| **Hourly Rate** | $ per hour |
| **Labor Cost** | Overall cost of the scheduled hours |
| **Status** | Schedule status |
| **Notes** | Manager comments |

**Employees WhenIWork Timesheet Table:**

|  |  |
| --- | --- |
| **Excused/Adjusted** | Manager manually overrides the attendance points incurred by the employee if justified. |
| **First Name** | First Name of the employee |
| **Last Name** | Last Name of the employee |
| **Employee ID** | EMPLID |
| **Date** | Shift Date |
| **Start Time** | Shift clock-in time |
| **End Time** | Shift clock-out time |
| **Unpaid Breaks** | Break time utilized |
| **Regular** | Clocked hours during the shift |
| **Hourly Rate** | $ per hour |
| **OT** | Overtime duration between 0 and 4 |
| **Double OT** | Double overtime duration between 4 and 14 |
| **Paid Total** | Total pay for the clocked hours including OT |
| **Schedule** | Schedule Name |
| **Job Site** | Location of the job |
| **Position** | Position at Learning Spaces |
| **Manager Note** | Managers note about the clocked hours |
| **Clock In Note** | Employees note while clocking in |
| **Clock Out Note** | Employees note while clocking out |

**Employees Time-off Requests Table:**

|  |  |
| --- | --- |
| **Excused/Adjusted** | Manager manually overrides the attendance points incurred by the employee if justified. |
| **Employee** | Employee full name |
| **Employee ID** | EMPLID |
| **Submitted By** | It can be either the employee themselves or the manager. |
| **Type** | Type of time-off request |
| **Created** | RTO requested date |
| **Start Date** | RTO start date |
| **Start Time** | RTO start time |
| **End Date** | RTO end date |
| **End Time** | RTO end time |
| **Paid Hours** | Paid hours if applicable |
| **Status** | Status of the request |
| **Approved/Denied By** | Manager Name |
| **Approved/Denied On** | Approval or Denial date |
| **Message** | Justification or comment for the time-off. |

**Employees Qualtrics Timesheet Correction Table:**

|  |  |
| --- | --- |
| **Team** | The team the employee work for: Learning Spaces or Services |
| **System** | The clock-in system name |
| **Timesheet Date** | Timesheet correction date |
| **Time In** | Correct clock in time |
| **Time Out** | Correct clock out time |
| **Location** | Location of the job |
| **Comments** | Comments from the employee |
| **Acknowledgement** | Acknowledging that the information is correct |
| **OU Email Address** | The email address of the employee |
| **First Name** | Employee first name |
| **Last Name** | Employee last name |
| **OU 4x4** | 4x4 of the employee |

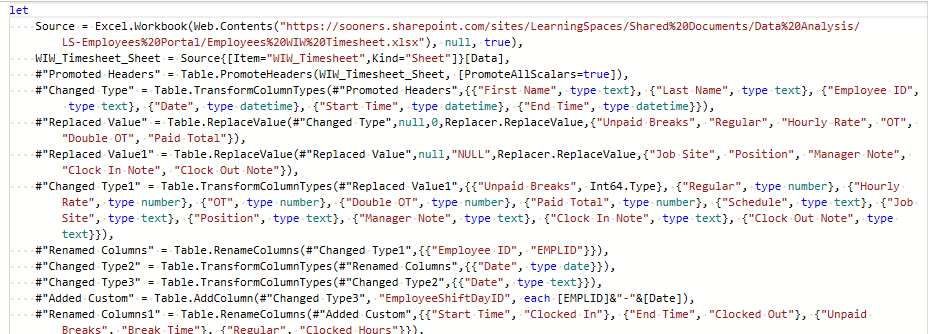
**User Access Table:**

|  |  |
| --- | --- |
| **First Name** | Employee first name |
| **Last Name** | Employee last name |
| **OU Email Address** | The email address of the employee |
| **OU 4x4** | 4x4 of the employee |
| **Sooner ID** | OU Net ID of the employee |
| **EMPLID** | The employee identification number |
| **Is Manager** | A flag to check if the employee given access as a manager or not |

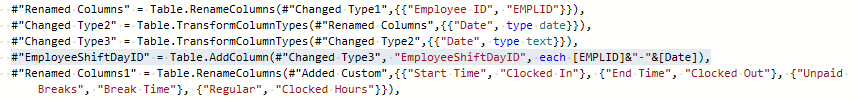
In addition to that, the manager is given the ability to update the dataset manually when the attendance infraction is justified by including a separate field that overrides the existing record.

1. **Data Preparation and Modelling:**

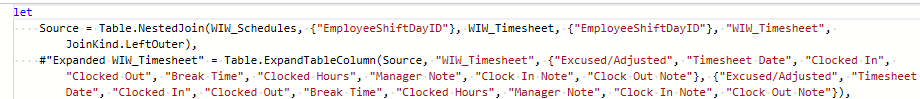
As the dataset is ready, it is fetched to Power BI and transformed to the desired shape by using Power Query as follows:



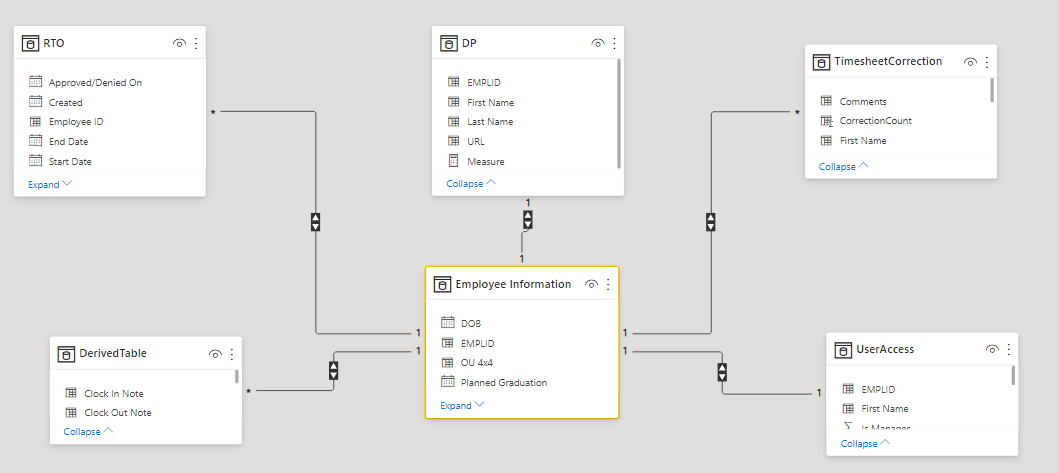
Since the tables: WIW\_Schedules and WIW\_Timesheet do not have a common primary key to uniquely identify an entity occurrence for joining operation so I created a composite key on each table by combining EMPLID and Shift Date and named it as EmployeeShiftDayID.



This new key is used to join both the tables by LEFT OUTER JOIN to capture no-show for the shift as there is no entry in WIW\_Timesheet table if the employee misses their shift.

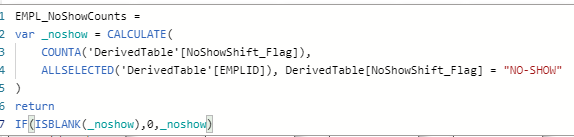


All these tables have EMPLID as either PRIMARY KEY or can act as a FOREIGN KEY so I have joined all these tables ensuring the data integrity and following the below cardinality:

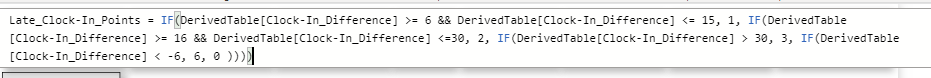


Following the above data integrity and cardinality, the data is loaded to the model. Since not all the columns are required for the visualization and analysis, many of the columns have duplicates so many of the columns are truncated before loading to the model. The following measures are created for giving aggregated values and some for better insights:

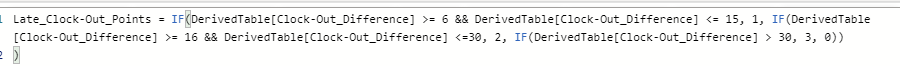
**For No-Show entries:** A measure to give the overall no-show count for the shifts missed.



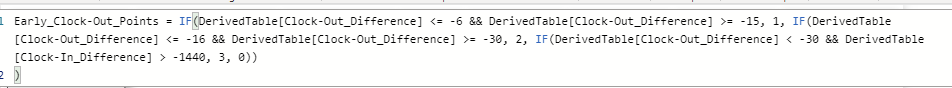
**Clock-in tardy points per shift:** Tardy points incurred by the employee when they clock in late.



**Clock-out late tardy points per shift:** Tardy points incurred by the employee when they clock out late.



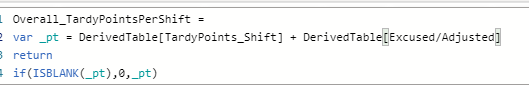
**Clock-out early tardy points per shift:** Tardy points incurred by the employee when they clock out early.



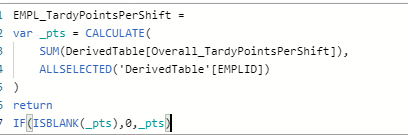
**Tardy points per shift (w/o justification):** The sum of late clock-in, early clock-out, and late clock-out points.



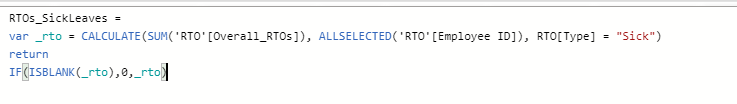
**Tardy points per shift (w/justification):** The overall tardy points when it’s justified by the employee and approved by the manager.



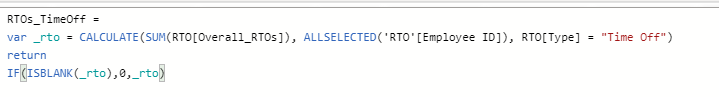
**Final Employee Tardy points**: A measure to display the overall tardy points per shift per employee



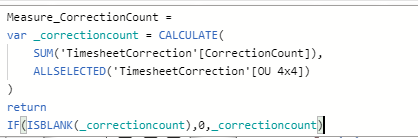
**Sick Leaves count:** A measure to give the overall count of Sick Leaves.



**Time-off requests:** A measure to give the overall count of time-off requests.



**Time correction requests**: A measure to give the overall count of timesheet correction requests.



Once these measures give the required insights, the model can be visualized by using the appropriate charts and graphs in the next step.

1. **Data Visualization:**

A mere presentation of data in a tabular form can serve little if not no purpose for the end user so I used the charts that are visually appealing and at the same time, giving insights to their logged data.

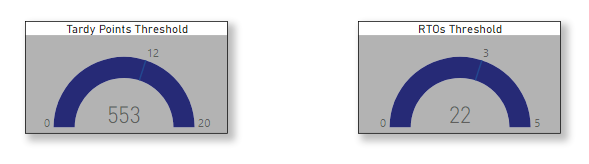
**Employee Information:** An updated information of the employee is required for the managers to follow-up and keep track of their graduation semester and plan new hires accordingly hence I used their individual information at the top.



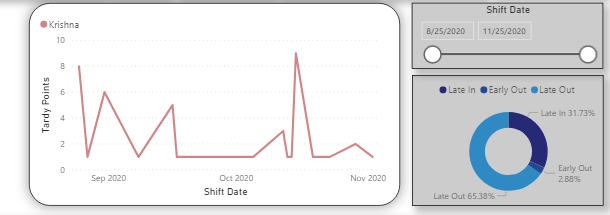
**For all the measures:** A card visual is used for all the measure to keep track of the tardy points, time-off requests, sick leaves, no-show, and timesheet correction requests.



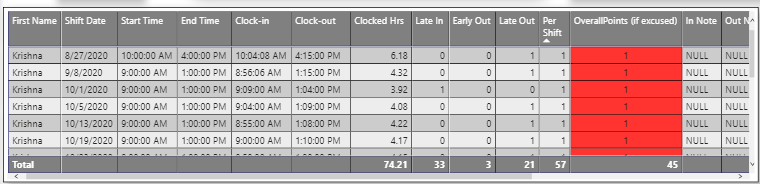
**Tardy Points and RTOs Guage:** The best way to represent the KPIs through gauge visual so I used this for tardy points and RTO counts to let the student employee know where they are leading as 12 tardy points lead to disciplinary action from the management.



**Tardy points trends:** The need to give a chance to look how tardy one is over the semester, can make much difference in mitigating the tardiness, so I included the line chart to give insights on incurred points. They can adjust the duration of the trend as well.



**Timesheet in tabular form:** The visual to have timesheet information with their schedule do serve the purpose of keeping track of their clocked hours and infractions for a much-detailed view.



1. **Data Evaluation:**

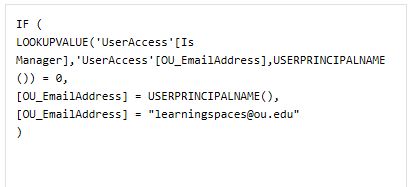
Once the visualizations are ready, a test plan is created to test all the columns and measures by using automated DAX and otherwise, manually. The following tests are done to ensure the data accuracy:

|  |  |  |
| --- | --- | --- |
| **Test Scenario** | **Test Case** | **Test Status** |
| Data Loading | Verify the data source pointed to Learning Spaces sharepoint folder. | **PASS** |
|  | Check if the Sharepoint and Power BI are connected well and there are no access issues. | **PASS** |
|  | Verify that all columns are loaded. | **PASS** |
|  | Verify that the record count is matched between Sharepoint file and Power BI Power Query. | **PASS** |
|  | Check for non-fatal errors. | **PASS** |
| Visualization | Display date and time. | **PASS** |
|  | Display ‘0’ instead of ‘(Blanks)’ for measures like Mean, Median, Standard Deviation | **PASS** |
|  | Compare the measures and the aggregated source data. | **PASS** |
|  | Check if the selection filters all the visuals | **PASS** |
|  | Check if the employee can see all the visuals when the app is launched. | **PASS** |
| Deployment | Check the report access permissions. | **PASS** |
|  | Verify if all the visuals are published to Service. | **PASS** |
|  | Check if the app is accessible only to Learning Spaces employees. | **PASS** |
|  | Check if the “Analyse to Excel” is disabled. | **PASS** |

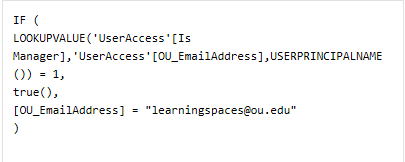
1. **Deployment and Publishing:**

The visualization and analytics part of the project was beta tested by student employees, and approved, and appreciated by the management. Finally, it was deployed to the cloud using Microsoft Service after ensuring the data authentication by coming up with Row-level data security at the dataset level, assigning the roles at the model level as follows:

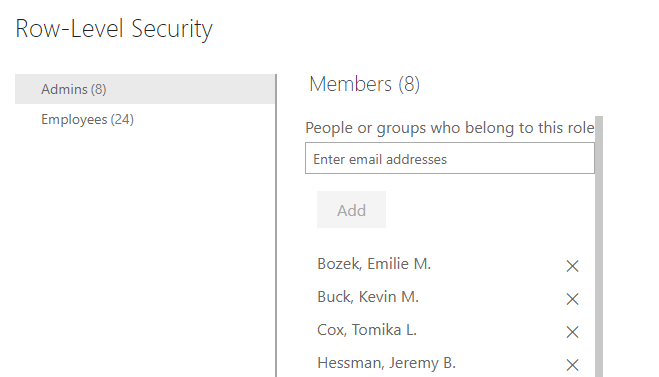
**For Employees Role:**



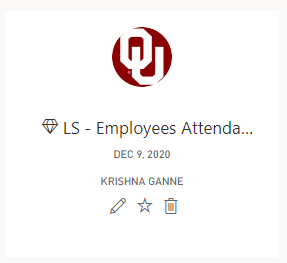
**For Manager Role:**



**Assigning roles on Service:**

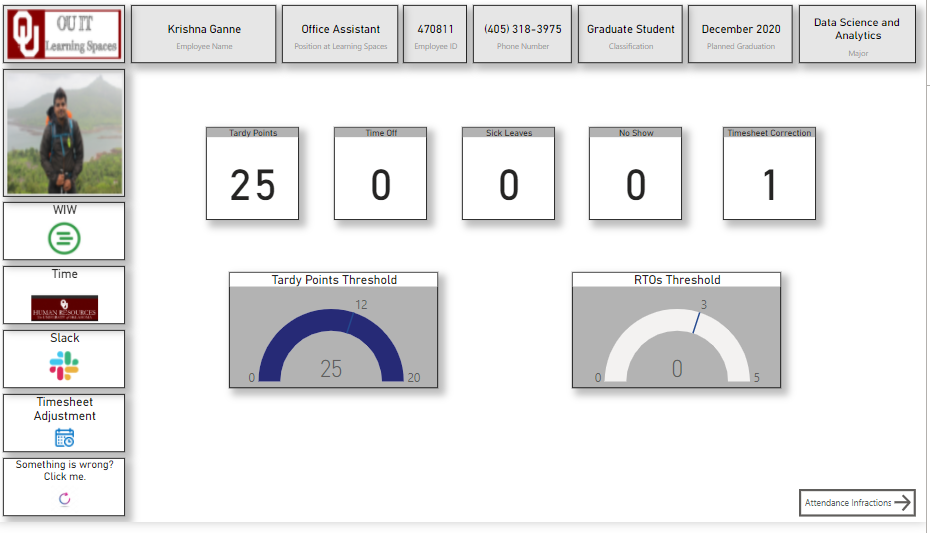


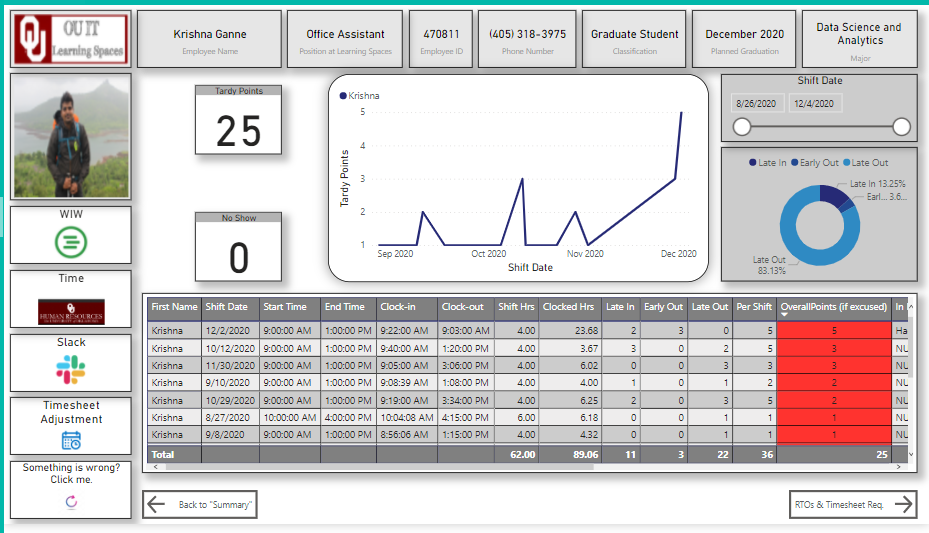
The portal is published, and an app is created for the same for better accessibility and to improve the user interface as follows.

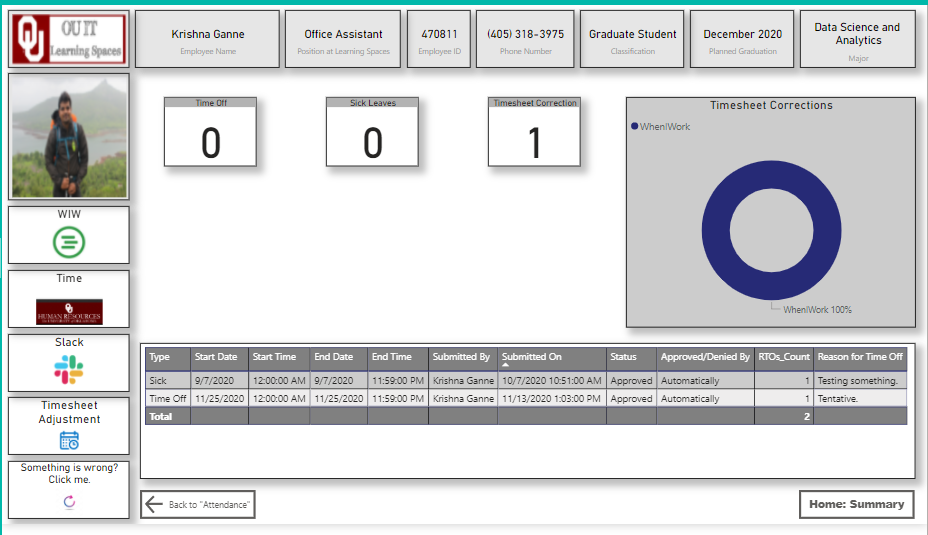


1. **Results:**

Student employees access the visualization through the link on private Sharepoint site or through an app, and the below visually appealing information pops up. Now the user can look at their and only their information adhering the privacy associated with timesheets, and attendance infractions by navigating through the pages.







**References:**

* OU IT Learning Spaces

( <https://www.ou.edu/ouit/learning> )

* Reza Rad,2019: *Power BI from Rookie to Rock Star, Book 4: Power BI Modeling and DAX*
* Microsoft Power BI Desktop Documentation

(<https://docs.microsoft.com/en-us/power-bi/fundamentals/desktop-getting-started>)

* Microsoft Power BI Service

([*https://docs.microsoft.com/en-us/power-bi/admin/service-admin-rls*](https://docs.microsoft.com/en-us/power-bi/admin/service-admin-rls)*)*

* Reza Rad Blog for Row-level Security:

*(*[*https://radacad.com/dynamic-row-level-security-in-power-bi-with-exclude-and-include-rules*](https://radacad.com/dynamic-row-level-security-in-power-bi-with-exclude-and-include-rules)*)*

* Microsoft Data Preparation Strategies

([*https://powerbi.microsoft.com/en-us/blog/tag/data-preparation/*](https://powerbi.microsoft.com/en-us/blog/tag/data-preparation/)*)*