|  |  |
| --- | --- |
| **Assignment Understanding and Visualizing Data** | |
| **Part 1: Data Collection Plan** | |
| **Step 1:** | |
| **What is the Situation?** | Data Center Wide Discovery and Assessment |
| **Who are the key stakeholders?** | |  |  |  | | --- | --- | --- | | **Stakeholder** | **What’s at Stake? If it goes wrong** | **What’s at stake? If it goes right** | | Retail Industry Infra Team | Failure in optimizing the infra | * Successful optimization of Infra. * Reduction in effort and cost | | Application Owner | Wrong Decision would impact the application and the Business stream using the same | Right decision is going to make sure there is no impact to the Application or Business | | Vendor Supporting the Discovery and Assessment | Reputation loss and loss of future business | Multiple Business Opportunity and Good Reputation | |
| **What are the parameters or options in the decision?** | Data we collect and Pivot Table which will be implemented and shared with stakeholders.  Insights of Business Owners related to their Application  Insights of Data Center Team |
| **How do you hope or expect data to help illuminate your decision?** | The Data we are planning to collect will help us in:   * Build Accurate Application Portfolio * Identify and Define Move Groups (Categorization and Future State) * Wave Planning (implementation Roadmap) |
| **Step 2:** | |
| **What are the key performance indicators for your situation?** | |  |  |  |  | | --- | --- | --- | --- | | Stakeholders  KPIs | *Retail Industry Infra team* | *Application Owner* | *Vendor Supporting the Discovery and Assessment* | | *Accurate Application Portfolio* | Yes | NA | Yes | | *Categorize the Value of application* | Yes | NA | Yes | | *Determine Future State* | Yes | Yes | Yes | | *Implementation Roadmap* | Yes | Yes | Yes | |
| **What defines the range of cases you will consider?** | The Range of Cases we will consider as part of this Discovery and Assessment are:  Environment type like (Development, Staging, Production)  Moving to Cloud or Not  In Cloud multiple Options like (IaaS, PaaS, SaaS)  If not moving to cloud, Options like (Decommission etc) |
| **What are the variables you will consider?**  **Is each variable categorical or quantitative?**  **What purpose does each variable have in informing your decision?** | |  |  |  | | --- | --- | --- | | Variable | Categorical/Quantitative | Purpose | | Host Name | Categorical | Identifier | | Environment | Categorical | Categorize into buckets | | Connection Count | Quantitative | Identify close affinity | | Connection Direction | Categorical | Identify affinity direction | | Remote Host Name | Categorical | Identify close affinity | |
| **Step 3:** | |
| **Where will the data come from?** | * We will install a agent which will run on all the computers in the data center for a maximum of 60 days and minimum 30 days. * Prepare Interview Questionnaire to get the right information from Application Owners, Users and Support Team * Collect data from Internal workshops, Information library Portals, Application Registries. |
| **Is it observational or experimental data?** | It is Observational data |
| **Who will Collect it?** | Data is collected via:  Tool  Interviews,  Workshops  Registries,  Portals |
| **How much data will you need (sample size)?** | Maximum 60 days to minimum 30 days of data |
| **How will you assure that it is representative of the population?** | Since we are collecting 2 months max of data. The pattern of usage would be stable and inline with standard usage. Hence, 2 months sample should be representative of the population |
| **What steps will you take to mitigate potential bias?** | Interviews, workshops, registries, portals information will be validated with 2 months of usage data. The summary of usage would give idea about any bias in the data collected manually. |
| **Part 2: Identify Data Summaries and Visualization** | |
|  | |
| **What summary statistics will you use to inform your decision?** | Max Connection Count per Host Name will give the insight |
| **Are you interested in statistics that are sensitive to or resistant to outliers, and why?** | Yes, We work on Connection Counts and all Connection Counts related to Monitoring Tools, Network Appliances, Duplicate Servers will be manually filtered to eliminate unnecessary outliers. |
| **Visualization** | Bar chart |
| **Part 3: Data and your decision** | |
| **Part a:** | |
| **What are you attempting to model with your data?** | Trying to get insights about the close relationship a compute has with other servers. |
| **What are the KPIs for the situation you are trying to understand?** | |  | | --- | | Accurate Application Portfolio | | Categorize the Value of application | | Determine Future State | | Implementation Roadmap | |
| **What is the relationship between your variables and the KPIs?** | Connection Count, Remote Host Name, Host Name will give insights about the relationship each other have using which KPI’s like Accurate Application Portfolio, Categorize the Value of application, Determine Future State, Implementation Roadmap can be determined easily |
| **What are the limitations of your model** | The model suffices the KPI requirement as such there is no limitations |
| **Do you feel your model is good enough** | Yes, as informed before the model suffices the KPI requirements and gives insights to take decisions seamlessly. |
| Part b: | |
| **Attached the Mock Document for Dashboard** |  |