# STATE EMPLOYEE CREDIT CARD TRANSACTION

## Data Analyst: The Hao Diep

## Client/Sponsor: Financial Department

## Purpose:

*To analyze state employee credit card transaction data in order to:*

1. *Detect potentially fraudulent or anomalous transactions (****KPI****: % high-value outliers, # flagged refunds)*
2. *Identify high-risk departments or merchant categories for targeted oversight (****KPI****: Top 5 in total spend, variance vs average)*
3. *Evaluate spending trends over time to inform cost management and budgeting (****KPI****: % increase/decrease per fiscal period, seasonal patterns)*

## Scope / Major Project Activities:

*What are the major parts of this project? List out the high-level steps, activities, or stages of the project, and give a brief description for each.*

|  |  |
| --- | --- |
| Activity | Description |
| Data Overview | Review and describe dataset structure, attributes, and data types |
| SMART Question Definition | Formulate SMART questions to guide the analysis |
| Data Cleaning | Identify and fix issues such as missing values, outliers, and inconsistent formats |
| Data Analysis | Use Python and Jupyter Notebook to summarize and explore the dataset |
| Data Visualization | Use Tableau to create dashboard with charts answering each SMART question |
| Modelling | Detect and explain outliers using Isolation Forest, LOF, and Z-score/IQR. |
| Insight Generation | Interpret findings and develop recommendations |
| Documentation | Compile results into report and upload project to Github |

## This project does not include:

* Development of real-time transaction monitoring systems
* Building predictive models for future spending beyond the scope of anomaly detection

## Deliverables:

*A specific list of things that your project will deliver.*

|  |  |
| --- | --- |
| Deliverable | Description/ Details |
| Cleaned Dataset | An Excel sheet with structured, cleaned data ready for analysis |
| Dashboard | Visual summary (5 charts) answering each SMART question |
| Insight Report | Key insights and recommendations derived from analysis |
| GitHub Repository | Project files (Excel, PDF report, README.md) organized and published online |

## Schedule Overview / Major Milestones:

*The expected schedule for the project. This can be defined by milestones (e.g. “all data is cleaned and processed”), periods of time (“Week 1 / Week 2”), or other ways based on the needs of the project.*

|  |  |  |
| --- | --- | --- |
| Milestone | Expected Completion Date | Description/Details |
| *Phase 1: Data Overview* | *August 4, 2025* | *Complete the data description and structure sheet* |
| *Phase 2: SMART + SOW* | *August 5, 2025* | *Identify questions, complete SOW file* |
| *Phase 3: Cleaning + Dashboard* | *August 7, 2025* | *Data processing, building summary tables, creating charts (3 dashboard with 3 charts for each)* |
| *Phase 4: Modelling* | *August 8,2025* | *Detect outliers using machine learning models (e.g., Isolation Forest)* |
| *Phase 5: Insight & Github* | *August 10, 2025* | *Extract insights, write reports, upload to GitHub* |

## \*Estimated date for completion:

*August 11, 2025*