CYBER RISK DASHBOARD

ILAN SARBAC DEVELOPER'S INSTITUTE DATE : 09/09/2025

Final Project – Cyber Risk Scoring & Dashboard

Developer's Institute – Data Analyst Bootcamp Student: Ilan Sarbac Date: September 2025

Project Overview

The goal of this final project is to simulate a realistic cyber risk scoring system for small and medium-sized enterprises (SMEs), based on enriched company data, and to visualize key insights using a professional dashboard. This project reflects a full analytical pipeline—from data preparation to business-oriented decision support.

Business Objective

Cyber insurance is becoming a critical need for companies, especially SMEs that are often underprepared but increasingly targeted by cyber threats. The objective was to:

- Identify which SMEs present the highest cyber risk
- Simulate a financial cost based on this risk
- Segment these companies for insurance underwriting or advisory purposes
- Visualize key patterns and trends in a dashboard

Tools & Technologies Used

Tool	Purpose
Google Colab	Data analysis & scoring logic in Python
Pandas / NumPy	Data processing and transformation
Matplotlib / Seaborn	Initial visualizations
Tableau (Mac)	Interactive final dashboard
CSV / Excel	Data export and sharing format

Technical Workflow

1. Data Import & Cleaning
   - Source file: companies_enriched.csv
   - Data quality check, null removal, type normalization
2. Feature Engineering
   - Created new column: Risk_Score (from multiple weighted factors)
   - Simulated estimated financial risk cost
   - Categorized companies by Risk_Level (Low / Medium / High)
   - Added a derived feature: Company_Size (based on employee count)
3. Scoring Formula (Python)

```
df['Risk_Score'] = (
    df['nb_employees'] * 0.1 +
    df['has_website'] * 5 +
    df['industry_risk_factor'] * 10 +
    df['email_exposure_score'] * 0.2 +
    df['tech_stack_risk_score'] * 0.5
)
```

1. Cost Simulation

```
df['Estimated_Risk_Cost'] = df['Risk_Score'] * 120
```

1. Risk Classification

```
def categorize_risk(score):
    if score < 150: return 'Low'
    elif score < 300: return 'Medium'
    else: return 'High'
```

```
df['Risk_Level'] = df['Risk_Score'].apply(categorize_risk)
```

1. Export to Tableau
    - Final dataset: companies_scored.csv
    - Used for dashboard creation and storytelling

Dashboard Insights (Tableau)

The final dashboard includes 5 key visualizations:

1. Pie Chart – Distribution by Risk Level → Shows how many SMEs fall into Low, Medium, or High risk categories.
2. Bar Chart – Average Estimated Risk Cost by Industry → Highlights which industries represent the highest financial exposure.
3. Stacked Bar – Risk Level by Company Size → Reveals whether small, medium, or large SMEs tend to be more exposed.
4. Donut Chart – Risk Level per Sector → Compares sectoral exposure, useful for underwriting or advisory segmentation.
5. Line Chart – Correlation Between Risk Score and Estimated Cost → Shows linear dependency and validates the scoring model logic.

Interpretation & Use Case

This simulated model allows for:

- Quick triage of SMEs based on cyber exposure
- Financial projection for risk-adjusted insurance pricing
- Visual decision-support tool for underwriters or cybersecurity consultants

It demonstrates how data science can empower real-world business decisions, especially in industries where risk quantification is essential.

What I Learned

- Building an end-to-end data pipeline, from raw CSV to insights
- Crafting a risk scoring algorithm using business logic and code
- Leveraging Tableau for business storytelling
- Translating analytical findings into actionable insights
- Documenting and communicating data science work clearly

Project Deliverables

Deliverable	Format
Python Simulator	.ipynb notebook on Google Colab
Cleaned Dataset	companies_scored.csv
Dashboard	Tableau (with 5 visuals)
Insights Summary	PDF / Loom Video Presentation

Final Notes

This project reflects my personal growth as a Data Analyst. It combines technical coding, visual clarity, and business acumen—applied to a very relevant and modern risk: cybersecurity. I hope it shows my ability to tackle real-world problems using data.