Ontry Patten

Southern New Hampshire University

CS-499 Enhancement 3\_Databases\_ Added the Data Protection Plan

DAT 220 Fundamentals of Data Mining



Bubba Gump Shrimp Company

Strategic Data Analysis & Protection Plan

Table of Contents

[Analysis Plan 3](#_Toc147052206)

[Introduction 3](#_Toc147052207)

[Business Problem 3](#_Toc147052208)

[Analytic Method 4](#_Toc147052209)

[**Analysis Process, Tools and Visualizations** 4](#_Toc147052210)

[Analysis Tools 4](#_Toc147052211)

[Data Visualizations 4](#_Toc147052212)

[Research Question 5](#_Toc147052213)

[Research Measurement 5](#_Toc147052214)

[Follow-Up Questions 5](#_Toc147052215)

[Research and Support 6](#_Toc147052216)

[Analysis 6](#_Toc147052217)

[**Preparing Data and Analysis** 6](#_Toc147052218)

[Analysis Organization 6](#_Toc147052219)

[Meaningful Patterns 9](#_Toc147052220)

[Inaccurate Depictions of Data 9](#_Toc147052221)

[Alternative Analytic Methods 10](#_Toc147052222)

[**Data Presentation and Report** 10](#_Toc147052223)

[Display and Interpretation 10](#_Toc147052224)

[Validity, Reliability, Limitations 11](#_Toc147052225)

[Resulting Decision Influence 12](#_Toc147052226)

[Visual Evaluation 13](#_Toc147052227)

[Next Steps 13](#_Toc147052228)

[APPENDIX A: Data Protection Plan (DPP) 14](#_Toc147052229)

[Purpose 14](#_Toc147052230)

[Critical Components & Vulnerabilities 14](#_Toc147052231)

[Vulnerabilities 16](#_Toc147052232)

[Zero Trust 18](#_Toc147052233)

[Database Security 20](#_Toc147052234)

[List of References 22](#_Toc147052235)

# **Analysis Plan**

## Introduction

The primary objective of this report is to perform a comprehensive analysis of the organization's data landscape with a particular focus on integrating robust cybersecurity measures. In an age where data is not only an invaluable asset but also a prime target for cyber threats, it is critical to evaluate our data assets, their vulnerabilities, and potential security risks. This analysis aims to identify strengths and weaknesses in our current data security posture, assess the effectiveness of existing cybersecurity measures, and propose enhancements to fortify our data defenses. By merging data analysis with a cybersecurity plan, we intend to establish a proactive and adaptive security framework that safeguards our critical information assets, ensures compliance with industry regulations, and enhances our overall resilience in the face of evolving cyber threats.

## Business Problem

Bubba Gump Shrimp Company experienced a surge in popularity following its mention in the blockbuster movie, Forrest Gump. Subsequently, they witnessed increased food and merchandise sales. However, the company's fame was short-lived, and profits started to plateau, with declining sales. In response, Bubba Gump Shrimp Company undertook data collection and analysis, including data from restaurants, online services, third-party retailers, Point of Sales (POS), online sales, and customer information. The leadership sought to better understand their customers and identify opportunities for new revenue growth through data analysis.

## **Analytic Method**

The strategy employed involves the analysis of various data sources provided by Bubba Gump Company, including Point of Sales (POS) data, customer loyalty programs, and online sales. Additionally, data from web sales, in-store sales, and third-party retailers are considered to provide a comprehensive view of customer behavior. This analysis aims to uncover customer habits and provide insights to shift the company's focus for attracting new business. Reports and data visualizations will be presented to the company's leadership to inform decision-making.

### **Analysis Process, Tools and Visualizations**

Analysis Tools

In this analysis, we will utilize data mining tools, specifically Microsoft Excel and the SNHU JMP Data Analysis Tool. Microsoft Excel's built-in formulas will aid in data analysis, while the JMP tool is well-suited for comprehensive data analysis and visualization. JMP offers a more efficient approach to analyzing all available data simultaneously, making it a well-rounded tool for this task.

Data Visualizations

To present the findings effectively, we will employ the Scale Lift Chart, which is a performance measurement tool. This chart will assist in evaluating the impact of changes in data over time, particularly in differentiating between older and newer data. Additionally, we will use other visualization techniques to highlight discrepancies and identify successful strategies for the company.

## Research Question

The initial research questions that need to be addressed include understanding the reasons behind declining customer patronage and the factors contributing to declining profits. A specific focus will be placed on examining the average age group of loyal customers, as this may reveal meaningful patterns in the data.

Research Measurement

To assess the effectiveness of the research, we will examine the trends related to customer ages and their correlation with in-store, web, and third-party purchases. This analysis will provide insights into potential adjustments needed to increase profits.

Follow-Up Questions

Follow-up questions will be valuable in building upon the initial research. For example, questions such as "Did the 500-customer survey collect the correct data?" and "Does individual income impact in-store purchases?" will be explored based on the outcomes of data analysis to help achieve the desired research objectives.

## Research and Support

Several online resources will support this project, including articles like "Data Analysis: What, How, and Why to Do Data Analysis for Your Organization" and "16 Data Mining Techniques: The Complete List." These resources will provide detailed information on data analysis techniques, tools, and the importance of conducting business data mining analysis.

**Data Analysis: What, How, and Why to Do Data Analysis for Your Organization**:

<https://www.import.io/post/business-data-analysis-what-how-why/>

**16 Data Mining Techniques: The Complete List**:

<https://www.talend.com/resources/data-mining-techniques/>

# Analysis

### **Preparing Data and Analysis**

## Analysis Organization

To analyze the Bubba Gump Shrimp data, I used the **Cluster analysis** to separate the customer population and placed them into small individual groups with similar qualities. Age, income, spend rate and marital status were factors in this analysis. I can also use this technique to separate the customers to help develop marketing ads and promotions. For example, the customer group for the amount spent at the restaurant can show how likely they can manage to pay online if they make a web purchase. Bubba Gump can offer gift certificates or customer loyalty points to make purchases online. However, I think there is a minor discrepancy in this analysis. I used a single data point to separate customers. It is recommended that the cluster analysis technique use more than one data point to be effective in separating customers.

With the**Linear regression** analysis, I calculated the customer's web store spending. To calculate web store spending, I used some of the variables. I decided to use only the variables that were useful. For example, the restaurant spend and the website visits. With just these two variables in the model, the model isn’t practical. I found more valuable results when I analyzed spend rates at other store types and then tried to build a different model for different customer groups found in the cluster analysis. Information about variable relationships in linear regression is limited by the variables included in the model.

**To suggest if a customer would make a web purchase, a Logistic regression can be** used to predict such action. For example, in Fig 1 below, I kept analyzing the data using logistic regression to calculate other categories. a logistical regression model can be used by comparing the number of visits. We can see that there is approximately a 50% chance a customer will make a purchase in one website visit. This can help the Bubba Gump Shrimp Company develop marketing strategies to make the customers come back often.

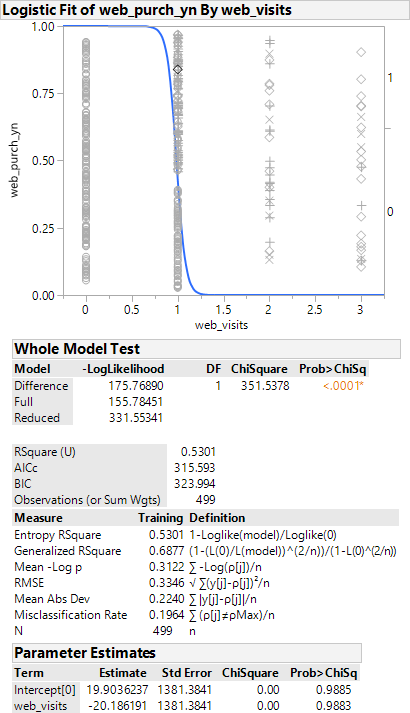


Fig. 1

**Sources of Error**

When looking at **Sources of Error** in the data, I noticed that there were customers spending with a value of zero for web spend. I could not properly verify which value was incorrect because some of the data was missing or was not in the correct format. For instance, the zip code was not in the correct format because there were some zip codes that started with a zero. To correct this issue, the zero needed to be added the front of all of the ones with a four-digit zip code. Once this was done, I will carefully analyze the dataset again with a stepwise approach because the data is a little skewed and may not give me an accurate report. In Fig 2 below you can see how the zip codes and missing the zeros. I will have to continue the process, because it’s better to have to collect additional data after the initial analysis is done.

Meaningful Patterns

In this case study, I don’t think there were any Meaningful Patterns or standouts. I will have to develop some additional questions to help capture meaning data. For instance, I will ask:

* Do customers who spend from third parties also share the same connection with restaurant spend rate?
* What is distinct difference of customers that spends at both web stores and restaurants?

## Inaccurate Depictions of Data

In this report, the Inaccurate Depictions of Data are mostly due to the extreme number of zeros found in the spend amount column. I felt like it really skewed the data results and may need to remove the zeros from the analysis.

Alternative Analytic Methods

Other Alternative Analytic Methods can include the following:

* Hierarchical Clustering
* K-means clustering
* Decision Trees

Each of the alternative methods has strengths and weaknesses. Let’s take The Decision Trees for example. The Decision Tree do not need any statements of linearity and could be useful with analyzing customer traits and variables. The Decision trees are easy to interpret and connect to the business.

### **Data Presentation and Report**

Display and Interpretation

Once again, we analyzed the Bubba Gump dataset. After another closer look, we can conclude that there is a relationship between the website spending/visits and the restaurant visits. The data results suggest that marketing efforts should be focused on customers who visit physical restaurant locations. We can also determine from this analysis that other restaurant locations had more customers visits. Because of the correlation between restaurant visits and website spending, we can change our focus and increase the research and analysis to website spending. I used the Cluster analysis because we didn’t know how many clusters existed in the data before running the model. The study showed that customers who will be most likely to shop on the Bubba Gump website are around their 30s and averaging 50k annually.

Cluster Analysis Web\_Sp end vs. Restaurants Spend

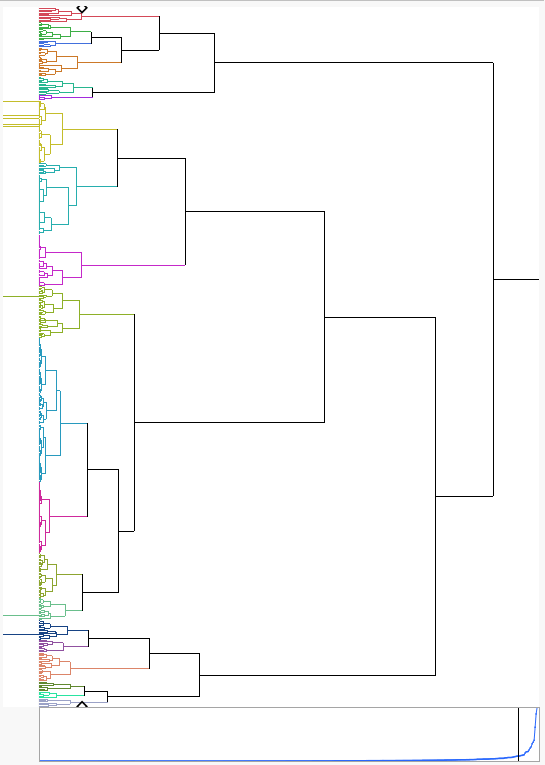


Fig. 2

## Validity, Reliability, Limitations

When analyzing data, we expect a certain number of errors and missing data. We are given the Bubba Gump Shrimp company data information in this exercise. After looking over the data, I noticed that some of the data were missing and hard to recognize. I noticed that some restaurant locations were not filled, and some places appeared to be left entirely out. To thoroughly analyze this part of the dataset, additional data would need to be collected so all locations would be adequately represented. When considering the reliability of the analysis, the validations steps must be accurate. I think applying control measures when data is collected can ensure minimal errors. Since the Bubba Gump dataset contains 500 responses, the limitations should be considered before the analysis begins. I would recommend a comparison against additional data from new customers.

## Resulting Decision Influence

Looking at the clustering analysis (Fig 1), we can see that it can influence some marketing efforts. For example, Bubba Gump Shrimp can focus on customers in the most profitable restaurant locations.

Bubba Gump Restaurants Visits vs. Restaurants Locations

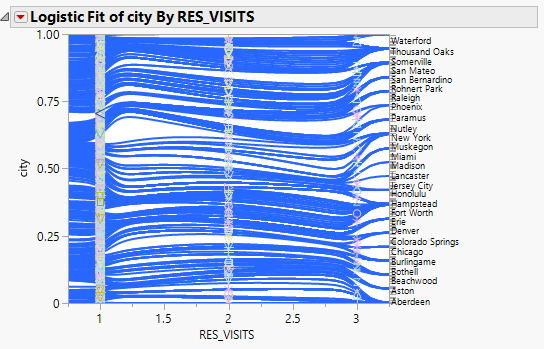


Fig. 3

Bubba Gump Restaurants visits by State

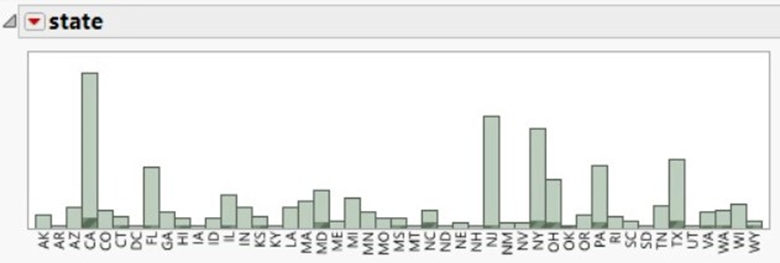


Fig. 4

Looking at Fig 2, we can see that some restaurants have loyal customers who repeatedly come to the same store. This can have a lot of influence on the decision-making of the how-to target customers.

Visual Evaluation

Cluster analysis charts can be difficult to interpret but visualizations instrumental in presenting the results in a readable manner. I prefer to use a combination chart of bars and lines because they are more readable. I will incorporate additional charts and add a summary in case the audience is not familiar with it.

## Next Steps

I think the next steps would strongly advise the client to pull more sample data from the data warehouse, so we could look more closely at restaurant spendings and gather a fuller picture for all locations. Even though the data that was provide, it seems as though it was lacking another venue. We can possibly take another a closer look at profits from third-party vendors against to help determine whether or not it is a beneficial venue for Bubba Gump. An alternative is probably eliminate this venue and only stay with store restaurants and webstore purchases.

# APPENDIX A: Data Protection Plan (DPP)

## Purpose

The Data Protection Plan (DPP) at Bubba Gump Shrimp serves the vital purpose of establishing a unified front for the security of our interconnected systems, valuable data, assets, and personnel. This plan functions as a comprehensive security reference accessible to all members of our organization. As data protection continually evolves in response to emerging threats and regulatory changes, the DPP ensures our adaptability and resilience. Its core objectives encompass cultivating a pervasive culture of data security, ensuring strict adherence to legal and regulatory mandates, and safeguarding the reputation of the Bubba Gump Shrimp brand. To guarantee the plan remains up-to-date and aligned with current industry standards and legal obligations, we strongly advocate for regular consultation with legal experts and cybersecurity professionals. The Data Protection Plan represents an integral pillar in fortifying our commitment to data security, thereby underpinning the enduring prosperity of Bubba Gump Shrimp.

## Critical Components & Vulnerabilities

The Critical database components are vulnerable to attacks due to several factors. These components, including database servers, data stores, and associated infrastructure. These items contain valuable and sensitive information, making them attractive targets for malicious actors. Vulnerabilities can arise from misconfigurations, unpatched software, weak access controls, or insider threats. These discrepancies create opportunities for unauthorized access, data breaches, and manipulation. Furthermore, the interconnected nature of modern databases, often accessed remotely or through web applications, expands the attack surface, increasing the risk of exploitation. To mitigate these vulnerabilities, it is crucial to implement robust security measures, adhere to the principle of least privilege, and continuously monitor and adapt to emerging threats. Implementing controls can safeguard critical data assets and maintain database integrity. The Table below is a list of critical components.

|  |  |  |
| --- | --- | --- |
| **Table 1. Critical Components List** | | |
| **Component** | **Significance** | **Security Measure** |
| Database | The database holds sensitive customer data, financial records, and proprietary recipes. A breach could lead to data loss, legal liabilities, and loss of customer trust. | Implement robust access controls, encryption for data at rest and in transit, regular database vulnerability assessments, and real-time monitoring for unusual activity. |
| Application and Web Servers | These servers host the restaurant's website and ordering systems. A breach could disrupt online operations, impact customer orders, and damage the company's online reputation. | Regularly update software and patches, deploy Web Application Firewalls (WAFs) to mitigate web-based attacks, and conduct penetration testing to identify |
| Business Process Management Software | This software streamlines restaurant operations. Breaches in this system could disrupt essential processes, impacting customer service and business efficiency. | Implement role-based access controls, audit logs for process changes, and ensure secure coding practices in custom applications. |
| Key System Interfaces | Interfaces connect various systems and enable data flow. Breaches in these interfaces could disrupt data integrity and compromise the flow of information. | Encrypt data transferred through interfaces, implement strong authentication mechanisms, and regularly monitor and test interfaces for vulnerabilities. |

To protect these critical components, consider the following security measures:

* **Regular Vulnerability Assessments:** Conduct routine assessments and penetration tests on all critical components to identify vulnerabilities and weaknesses.
* **Access Control:** Implement strict access controls based on the principle of least privilege to ensure that only authorized personnel can access critical systems and data.
* **Encryption:** Utilize encryption for sensitive data both in transit and at rest, ensuring data remains secure even if systems are compromised.
* **Intrusion Detection and Prevention:** Deploy intrusion detection and prevention systems to detect and mitigate security threats in real time.
* **Patch Management:** Maintain a proactive patch management process to promptly address security vulnerabilities in operating systems, software, and applications.
* **Incident Response Plan:** Develop a comprehensive incident response plan to address security incidents promptly to minimize their impact on critical components.
* **Employee Training:** Train employees on security best practices and ensure awareness of the importance of safeguarding critical components.

These security measures for the critical components, can strengthen its protection against potential threats and minimize the risks associated with data breaches or system compromises, thus preserving its reputation and business continuity.

## Vulnerabilities

A vulnerability is a weakness or flaw in the security of a system, process, or network that can be exploited by malicious actors to gain unauthorized access, disrupt operations, or compromise data integrity. For Bubba Gump Shrimp, vulnerabilities pose a significant threat as they can lead to severe consequences. Exploited vulnerabilities could result in data breaches, exposing sensitive customer information, proprietary recipes, and financial records. Such incidents could lead to reputational damage, legal liabilities, and financial losses.

| **Table 2. Vulnerabilities** | | |
| --- | --- | --- |
| **Threat** | **Description** | **Consequence of threat realization** |
| Command Injection | Injection flaws, such as SQL, OS, and DLAP injection occur when untrusted data is sent to an interpreter as part of a command or query | Command Injection can result in data loss or corruption, lack of accountability, denial of access, or complete host takeover |
| Cross Site Scripting (XSS) | XSS flaws occur whenever an application takes untrusted data and sends it to a web browser without proper validation or escaping | XSS allows attackers to execute scripts in the victim’s browser, which can hijack user sessions or redirect the user to malicious sites |
| Insecure Direct Object References | A direct object reference occurs when a developer exposes a reference to an internal implementation object, such as a file, directory, or database key. | Can compromise all data that can be referenced by the parameter. If object references are predictable, attackers can gain access to all available data of that type |
| Cross-Site Request Forgery (CSRF) | A CSRF attack forces a logged-on victim’s browser to send a forged HTTP request, including the victim’s session cookie and any other automatically included authentication information, to a vulnerable web application | Allows the attacker to force the victim’s browser to generate requests the vulnerable application thinks are legitimate requests from the victim. Can cause the system to perform any stat-changing operation the victim is authorized to perform |
| Denial of Service Attack (DoS) | A DoS attack is an attempt to make a resource unavailable to its intended users by manipulating network packets, flooding the network with useless data. | System unavailability - users are unable to access the targeted resource |
| Elevation of Privilege | Exploitation of a bug, design flaw, or configuration oversight in the operating system, database, or application software to gain elevated access to resources that are intended to be protected | Attackers access unauthorized functionality. Administrative functions are often key targets for this type of attack. |

To react effectively to vulnerabilities, Bubba Gump Shrimp should adopt a proactive approach. This includes regular vulnerability assessments to identify weaknesses, prompt patching and updating of software and systems, and the implementation of a intrusion detection systems. Additionally, having an incident response plan (IRP) in place allows for swift action if a vulnerability is exploited, helping to minimize the potential harm.

## Zero Trust

Zero Trust is a cybersecurity approach that challenges the traditional security perimeter-based mindset. It operates under the assumption that no one, whether inside or outside the organization, should be trusted by default. Instead, it requires continuous verification of the identity and security posture of anyone or anything trying to access resources on a network. today's cybersecurity landscape, where traditional perimeter defenses are increasingly less effective in protecting against advanced threats and where remote work and cloud services have expanded the attack surface. Implementing Zero Trust principles can help organizations improve their security posture by assuming that threats can come from both external and internal sources and requiring constant verification and security controls to mitigate these risks. Key principles and components of the Zero Trust model include:

* **Verification and Authentication:** Users and devices must be authenticated and verified before they are granted access to resources. This goes beyond just username and password, often incorporating multi-factor authentication (MFA) and other advanced identity verification methods.
* **Micro-Segmentation:** Networks are divided into smaller segments, and access is restricted to only those segments and resources necessary for a user or device to perform their specific tasks. This limits lateral movement within the network if an attacker gains access.
* **Least Privilege Access:** Users and devices are given the least amount of access necessary to perform their job functions. This ensures that even if an account is compromised, the potential damage is limited. Failure to adhere to the Principle of Least Privilege can have severe consequences for an organization. In the table below, is an example matrix tool used to ensure that no single individual or entity has too much control or influence over critical processes or functions within an organization:

|  |  |  |
| --- | --- | --- |
| Table 3. Separation of Duty (SoD) Matrix | | |
| **Access/Authorization** | **Responsible Individual(s)** | **Role Description** |
| Data Access | IT Administrator | Responsible for controlling access to data but should not have authority over payment authorization or data entry. |
| Data Entry | Accounting Team | Handles data entry but should not have control over system configuration or vendor management. |
| Payment Authorization | Finance Manager | Has the authority for payment authorization but should not perform database administration. |
| Database Administration | Database Administrator | Responsible for database administration but should not have control over system configuration or payment authorization. |
| System Configuration | System Administrator | Manages system configuration but does not handle data entry or payment authorization. |
| Vendor Management | Procurement Team | Manages vendor relationships but does not have access to sensitive data or perform database administration. |

* **Data Masking and Redaction:** Implement data masking and redaction to limit the exposure of sensitive data. This ensures that only authorized users see the complete data, while others see only masked or redacted information.
* **Data Backup and Recovery:** Regularly back up database data and ensure the backups are stored securely and separately from the primary database.

Applying these principles to the database and it’s components can significantly enhance its security posture and better protect sensitive data from a variety of potential threats, whether they originate from internal or external sources.

## Database Security

Securing a database involves designing a network topology that minimizes attack vectors, enforces access controls, and monitors for suspicious activities. Below is diagram of the Bubba Gump server network.

A diagram of a computer network

Description automatically generated

Figure 5. Bubba Gump Database Network Topology

This network topology and security measures (Table 2.) creates multiple layers of defense, helping to secure the critical database components and reduce the risk of successful attacks.

\*Note: Security is an ongoing process that requires vigilance and adaptation to emerging threats and vulnerabilities.

# List of References

Ahlemeye-Stubbe, A. & Coleman, S. 2014. A Practical Guide to Data Mining for Business and Industry. John Wiley & Sons, Ltd.

Earley, S. 2017. The problem of personalization: AI-driven analytics at scale. *IT Professional, 19*(6), 74-80.

Harvey, C., 2021. What Is Data Analytics? A Guide to Data Insights | eWEEK. [online] eWEEK. Available at: <https://www.eweek.com/big-data-and-analytics/data-analytics/>

Predictive Analysis Today Review Desk. (n.d.). JMP. Review: JMP in 2020. Retrieved from https://[www.predictiveanalyticstoday.com/jmp/](http://www.predictiveanalyticstoday.com/jmp/)

Rizzo, T. (2015). Maximizing results with data-driven financial marketing strategies. *Journal of Brand Strategy, 4*(2), 114-126.

06, B. P. | J. (2020, January 6). Why the zero trust model will become the new security norm. ITPro Today: IT News, How-Tos, Trends, Case Studies, Career Tips, More. https://www.itprotoday.com/cloud-security/why-zero-trust-model-will-become-new-security-norm