**Technical Appendix**

**Catch the Pink Flamingo Analysis**

**Produced by: <Fill in your Name>**

Acquiring, Exploring and Preparing the Data

Data Exploration

Data Set Overview

The table below lists each of the files available for analysis with a short description of what is found in each one.

|  |  |  |
| --- | --- | --- |
| **File Name** | **Description** | **Fields** |
| <Fill In> | <Fill in short phrase> | <Fill In: Name and describe all fields> |
| <Fill In> | <Fill in short phrase> | <Fill In: Name and describe all fields> |
| <Fill In> | <Fill in short phrase> | <Fill In: Name and describe all fields> |
| <Fill In> | <Fill in short phrase> | <Fill In: Name and describe all fields> |
| <Fill In> | <Fill in short phrase> | <Fill In: Name and describe all fields> |
| <Fill In> | <Fill in short phrase> | <Fill In: Name and describe all fields> |
| <Fill In> | <Fill in short phrase> | <Fill In: Name and describe all fields> |
| <Fill In> | <Fill in short phrase> | <Fill In: Name and describe all fields> |
| <Fill In> | <Fill in short phrase> | <Fill In: Name and describe all fields> |

Aggregation

|  |  |
| --- | --- |
| Amount spent buying items | <Fill In> |
| # Unique items available to be purchased | <Fill In> |

A histogram showing how many times each item is purchased:

<Fill In - upload screenshot>

A histogram showing how much money was made from each item:

<Fill In - upload screenshot>

Filtering

A histogram showing how many times each category of advertisement was clicked-on:

<Fill In - upload screenshot>

The following table shows the total amount of ad-click revenue for a set of specific values based on the advertisement category. All non-listed categories generate .25 revenue.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Scenario #** | **Electronics** | **Fashion** | **Automotive** | **Total Revenue** |
| 1 - even | 0.50 | 0.50 | 0.50 | <Fill In> |
| 2 - uneven | 0.55 | 0.60 | 0.55 | <Fill In> |

Data Classification Analysis

**Data Preparation**

Analysis of combined\_data.csv

Sample Selection

|  |  |
| --- | --- |
| **Item** | **Amount** |
| # of Samples | <Fill in> |
| # of Samples with Purchases | <Fill in> |

Attribute Creation

A new categorial attribute was created to enable analysis of players as broken into 2 categories (HighRollers and PennyPinchers). A screenshot of the attribute follows:

<Fill In Screenshot>

<Fill In: Describe the design of your attribute in 1-3 sentences.>

The creation of this new categorical attribute was necessary because <Fill in 1-2 sentences>.

Attribute Selection

The following attributes were filtered from the dataset for the following reasons:

|  |  |
| --- | --- |
| **Attribute** | **Rationale for Filtering** |
| <Fill in> | <Fill in 1-3 sentences> |
| <Optional Fill in> | <Optional Fill in 1-3 sentences> |
| <Optional Fill in> | <Optional Fill in 1-3 sentences> |
| <Optional Fill in> | <Optional Fill in 1-3 sentences> |

**Data Partitioning and Modeling**

The data was partitioned into train and test datasets.

The <Fill In> data set was used to create the decision tree model.

The trained model was then applied to the <Fill In> dataset.

This is important because… <Fill Nn>

When partitioning the data using sampling, it is important to set the random seed was set because… <Fill In>

A screenshot of the resulting decision tree can be seen below:

<Fill In>

**Evaluation**

A screenshot of the confusion matrix can be seen below:

<Fill In screenshot>

As seen in the screenshot above, the overall accuracy of the model is <Fill In>

<Fill In: Write one sentence for each of the values of the confusion matrix indicating what has been correctly or incorrectly predicted.>

**Analysis Conclusions**

The final KNIME workflow is shown below:

<Fill In screenshot>

What makes a HighRoller?

<Fill In 2-3 sentences answering this question based on insights from your analysis.>

|  |
| --- |
| **Specific Recommendations to Increase Revenue** |
| 1. <Fill In 1-2 sentences describing 1 recommended action> |
| 2. <Fill In 1-2 sentences describing a 2nd recommended action> |

Clustering Analysis

**Attribute Selection**

|  |  |
| --- | --- |
| **Attribute** | **Rationale for Selection** |
| <Fill in> | <Fill in 1-3 sentences> |
| <Optional Fill in> | <Optional Fill in 1-3 sentences> |
| <Optional Fill in> | <Optional Fill in 1-3 sentences> |
| <Optional Fill in> | <Optional Fill in 1-3 sentences> |

**Training Data Set Creation**

The training data set used for this analysis is shown below (first 5 lines): <Fill In Screenshot>

Dimensions of the training data set (rows x columns) : <Fill In>

# of clusters created: <Fill In>

**Cluster Centers**

|  |  |
| --- | --- |
| **Cluster #** | **Cluster Center** |
| <Fill In> | <Fill In> |
| <Fill In> | <Fill In> |
| <Optional Fill In> | <Optional Fill In>  (Note: add more rows to this table if you selected more than 3 clusters) |

These clusters can be differentiated from each other as follows:

Cluster 1 is different from the others in that… <Fill in 1-3 sentences>

Cluster 2 is different from the others in that… <Fill in 1-3 sentences>

<Optional Fill In> Cluster 3 is different from the others in that… <Fill in 1-3 sentences>

Note: Copy and fill in if you selected more than 3 clusters.

**Recommended Actions**

|  |  |
| --- | --- |
| **Action Recommended** | **Rationale for the action** |
| <Fill in 1-3 sentences> | <Fill in 1-3 sentences> |
| <Fill in 1-3 sentences> | <Fill in 1-3 sentences> |
| <Optional Fill in> | <Optional Fill in 1-3 sentences> |
| <Optional Fill in> | <Optional Fill in 1-3 sentences> |

Graph Analytics Analysis

**Modeling Chat Data using a Graph Data Model**

(Describe the graph model for chats in a few sentences. Try to be clear and complete.)

**Creation of the Graph Database for Chats**

Describe the steps you took for creating the graph database.

**Finding the longest conversation chain and its participants**

Report the results including the length of the conversation (path length) and how many unique users were part of the conversation chain. Include an image of the graph with the longest conversation chain.

<Insert image here>

**Analyzing the relationship between top 10 chattiest users and top 10 chattiest teams**

Include your table containing the top 3 chattiest users and teams below, and report whether or not any of the chattiest users are part of any of the chattiest teams.

**Chattiest Users**

|  |  |
| --- | --- |
| **Users** | **Number of Chats** |
|  |  |
|  |  |
|  |  |

**Chattiest Teams**

|  |  |
| --- | --- |
| **Teams** | **Number of Chats** |
|  |  |
|  |  |
|  |  |

**How Active Are Groups of Users?**

Report the top 3 most active users in the table below.

**Most Active Users (based on Cluster Coefficients)**

|  |  |
| --- | --- |
| **User ID** | **Coefficient** |
|  |  |
|  |  |
|  |  |

**Recommended Actions**

Finally, make recommendations to Eglence, Inc. and include examples of how your findings support them. Include this information in Slide 6 of your final presentation.