**Comprehensive Data Analysis Report: Bandcamp Sales Data**

**Executive** **Summary**

This report presents a comprehensive analysis of Bandcamp sales data between 9/9/2020 and 10/2/2020, covering data cleaning, exploratory data analysis and key business insights.

The dataset contains transaction records from the Bandcamp platform, including information about artists, albums, tracks, merchandise and additional metrics.

**1. Data Exploration and Cleaning**

**1.1 Initial Data Assessment**

The analysis began with an exploration of the database structure and the `bandcamp\_data` table specifically. Key columns identified included:

- Transaction identifiers (id)

- Artist and album information

- Item types (digital tracks, albums, physical items)

- Pricing and payment data

- Geographic information (country)

- Timestamps (‘utc\_date’)

Initial exploration revealed the need for data type standardization, null value handling, and column restructuring.

**1.2 Data Cleaning Process**

Significant data cleaning was performed to ensure data quality:

**1. ID Column Standardization:**

- Removed URLs from ID values

- Eliminated whitespace between numbers

**2. Currency Formatting:**

- Cleaned special characters from ‘amount\_over\_fmt’ column

- Standardized all monetary values to USD

**3. Categorical Value Standardization:**

- Transformed cryptic ‘item\_type’ codes ('a', 'b') to descriptive values

- Renamed ‘slug\_type’ values to meaningful categories

**4. Date Conversion:**

- Converted UNIX timestamps to standard DATE format

**5. Missing Value Handling:**

- Replaced empty artist names with "Unknown Artist"

- Set empty album titles to NULL

- Handled empty ‘track\_album\_slug\_text’ values

**6. Column Rationalization:**

- Removed duplicate and irrelevant columns

- Renamed columns for clarity (e.g., "\_id" to "id")

- Consolidated related columns

**2. Key Business Metrics**

The analysis revealed the following key metrics:



**3. Geographic Analysis**

**3.1 Top Countries by Sales Volume**

The top 15 countries by sales volume are:



**3.2 Revenue by Country**

The highest revenue-generating countries are:



**3.3 Customer Generosity by Country**

Countries with the most generous customers ("Die Hard" fans who pay significantly above asking price- that is, above the minimum price as sold by the artist):



**4. Product Analysis**

**4.1 Sales by Product Type**



**4.2 Pricing Analysis**

- Average price for digital items: $7.79

- Average price for physical items: $21.83

- Highest average price by artist: D.M.T Records at $952.51

**5. Artist Performance**

**5.1 Top Performing Artists**



**5.2 Artist Fan Engagement**

Artists with the most dedicated fans ("Die Hard" category):



**6. Customer Behavior Insights**

**6.1 Customer Segmentation**

Customers were segmented by generosity:

- Regular: Paid up to $50 above asking price

- VIP: Paid $50-$100 above asking price

- Die Hard: Paid over $100 above asking price

**6.2 Additional Purchases**

Analysis of additional music shows:

- 21742 transactions included additional music purchases

- Top artists for additional purchases: SAULT

**8. Digital vs Physical Media Comparison**



**9. Recommendations**

Based on the analysis, we recommend:

**1. Artist Engagement:**

- Focus marketing efforts on top-performing artists.

- Develop programs to reward artists with highly engaged fan bases.

- Provide top-performing artists with additional promotion tools.

**2. Geographic Targeting:**

- Increase promotion in high-value countries.

- Consider localized marketing for top territories.

**3. Product Strategy:**

- Expand merchandise offerings given their strong performance.

- Highlight merchandise to customers who previously purchased digital items.

- Bundle digital tracks to increase average order value.

- Target Gen-Z when marketing physical media as they are considered the nostalgic generation.

**4. Customer Loyalty:**

- Develop tailored VIP programs for highly engaged customers.

- Create special offers for customers who frequently purchase additional items.

**5. Data Collection Improvements:**

- Standardize data collection at source to reduce cleaning effort.

- Capture more detailed customer demographic information.

**10. Conclusion**

This comprehensive analysis of Bandcamp sales data has revealed valuable insights about customer behavior, product performance, and geographic trends. The findings can inform strategic decisions to optimize sales, improve customer engagement, and maximize revenue opportunities across different segments of the Bandcamp marketplace.

The creation of the `bandcamp\_sales` view consolidates these insights into a reusable resource for ongoing reporting and analysis needs.

**Appendix: Technical Details**

**Analysis Methodology**:

* Data cleaning using SQL functions and table alterations
* Aggregation and statistical analysis at multiple dimensions
* Time series analysis with weekly granularity
* Customer segmentation based on spending behavior

**Limitations**:

* Analysis limited to available data fields
* Time period constrained by dataset coverage
* Currency conversion based on historical rates