Digital Publishing Platform Analytics

User Engagement & Content Strategy Analysis

Engagement Analysis

Content Strategy

Tools: Python • Pandas • Statsmodels • Matplotlib

Presented by Thandiwe Moyo

Business Challenge

The Problem

Publishing platform needs to optimize user engagement and content strategy but lacks data-driven insights

User Engagement

Does session frequency correlate with deeper content consumption?

Content Distribution

Where are content gaps across languages and genres?

Business Impact

- **Revenue:** Ad optimization and subscription conversion
- **Strategy:** Content acquisition and localization priorities
- **Growth:** Audience targeting and retention improvements

Analytical Approach

Part 1: Engagement Analysis

- **Data Preparation:** Imported Google Analytics data, converted to minutes

- **Statistical Modeling:** OLS Linear Regression

- **Visualization:** Scatter plots with regression lines

Part 2: Content Analysis

- **Language Distribution:** Grouped titles across 7 languages

- **Genre Frequency:** Extracted 18+ genres

- **Export & Reporting:** CSV outputs for Business Intelligence (BI) tools

Key Research Question

"Does higher session engagement frequency lead to longer average engagement time, and where should we focus content investments?"

Technical Implementation

Regression Analysis Setup

Independent Variable: Engaged sessions per active user

Dependent Variable: Average engagement time (minutes)

Model: Ordinary Least Squares (OLS)

Data Cleaning Steps

- Time conversion (sec → min)
- Descriptive statistics review
- Multi-genre field parsing
- Label standardization

Tools & Libraries

- **Pandas:** Data manipulation
- **Statsmodels:** Regression
- **Matplotlib:** Visualization
- **NumPy & SciPy:** Statistics

Key Findings & Insights

Primary Finding

Positive relationship confirmed between engaged sessions per user and average engagement time

What This Means

Users with multiple engaged sessions show deeper content consumption patterns

Key: Session frequency indicates content value

Business Implications

- Prioritize retention over acquisition
- Invest in content discoverability
- Target high-frequency segments

Content Distribution Insights

Languages: 7 languages identified with expansion opportunities • **Genres:** 18+ genres mapped with content gaps

Strategic Recommendations

Retention Focus
Features encouraging repeat visits

Content Discovery
Improve recommendations

Market Expansion
Target underserved languages

Prioritized Action Items

HIGH **Launch retention campaign for high-frequency users**
Expected impact: +15-25% engagement time

MED **Diversify content in underrepresented genres**
Expected impact: +10-15% audience expansion

MED **Pilot localization in identified language gaps**
Expected impact: New market entry

Skills Demonstrated & Impact

Technical Skills

- **Statistical Analysis:** OLS regression
- **Data Manipulation:** Pandas
- **Visualization:** Matplotlib
- **Data Cleaning:** Text parsing

Business Skills

- **Strategic Thinking:** Revenue focus
- **Insight Translation:** Tech to action
- **Stakeholder Focus:** Executive outputs
- **Domain Knowledge:** Publishing

Measurable Business Impact

2	3	100%
Analytics Streams	Actionable Insights	Data-Driven

Project Outcomes

- Validated engagement strategy through statistical modeling
- Identified content gaps across languages and genres
- Delivered actionable recommendations