A Factor Analysis Approach to Persona Development using Survey Data

HAE MIN KIM
DATA ANALYST FOR RESOURCE ACCESS
HK433@DREXEL.EDU
ORCID.ORG/0000-0002-9128-0495

JOHN WIGGINS
DIRECTOR
LIBRARY SERVICES & QUALITY IMPROVEMENT
JWW27@DREXEL.EDU
ORCID.ORG/0000-0001-8617-5247
Personas

- Characters representing user groups
  - How they interact with products and services
  - To represent major user interests

- Our uses of personas
  - Predict generalized user behaviors
  - Help to share library stories with stakeholders

- Challenges
  - Qualitative data can take a long time to collect and interpret
  - Potential for researcher bias in interpreting results
Purpose

- To explore quantitative approaches
  - As a starting point
  - Application to survey data
  - To support part of library business cases
  - Using data including Drexel population as many as possible

- To identify major services and develop personas
  - Marketing plans
  - Business cases
  - Validation of developed personas
  - Web-user experiences
  - Physical space
Data Collection Methods

Data Collections

- Qualitative
  - Interviews
  - Focus groups
  - Observation
- Quantitative
  - Surveys
  - Record review
Method

- Survey
  - Questions about 15 library services
    - i.e. “How important is the library space to you?”
  - Data Collection
    - February and March in 2016
    - Distributed through
      - The Libraries website
      - Email via departments and student organizations
      - Facebook
      - Physical copies at the Libraries and classes
    - Partnership among library staff, business students and faculty
### Collected Data

- **Valid responses:** 435

### Demographic

<table>
<thead>
<tr>
<th>College</th>
<th>Enrollment</th>
<th>Percent</th>
<th>Responses</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arts and Sciences</td>
<td>3005</td>
<td>11.40%</td>
<td>48</td>
<td>11.10%</td>
</tr>
<tr>
<td>Biomedical Engineering Sci &amp; Health Systems</td>
<td>863</td>
<td>3.30%</td>
<td>29</td>
<td>6.70%</td>
</tr>
<tr>
<td>Biomedical Sciences &amp; Professional Studies</td>
<td>939</td>
<td>3.60%</td>
<td>21</td>
<td>4.80%</td>
</tr>
<tr>
<td>Business</td>
<td>3898</td>
<td>14.80%</td>
<td>39</td>
<td>9.00%</td>
</tr>
<tr>
<td>Computing and Informatics</td>
<td>1818</td>
<td>6.90%</td>
<td>13</td>
<td>3.00%</td>
</tr>
<tr>
<td>Economics</td>
<td>253</td>
<td>1.00%</td>
<td>2</td>
<td>0.50%</td>
</tr>
<tr>
<td>Education</td>
<td>1112</td>
<td>4.20%</td>
<td>20</td>
<td>4.60%</td>
</tr>
<tr>
<td>Engineering</td>
<td>4649</td>
<td>17.60%</td>
<td>24</td>
<td>5.50%</td>
</tr>
<tr>
<td>Entrepreneurship</td>
<td>13</td>
<td>0.05%</td>
<td>2</td>
<td>0.50%</td>
</tr>
<tr>
<td>Hospitality and Sport Management</td>
<td>409</td>
<td>1.60%</td>
<td>28</td>
<td>6.40%</td>
</tr>
<tr>
<td>Law</td>
<td>441</td>
<td>1.70%</td>
<td>5</td>
<td>1.10%</td>
</tr>
<tr>
<td>Media and Arts Design</td>
<td>2083</td>
<td>7.90%</td>
<td>55</td>
<td>12.60%</td>
</tr>
<tr>
<td>Medicine</td>
<td>1083</td>
<td>4.10%</td>
<td>73</td>
<td>16.80%</td>
</tr>
<tr>
<td>Nursing and Health Professions</td>
<td>4931</td>
<td>18.70%</td>
<td>48</td>
<td>11.00%</td>
</tr>
<tr>
<td>Honors</td>
<td>28</td>
<td>0.11%</td>
<td>0</td>
<td>0.00%</td>
</tr>
<tr>
<td>Professional Studies</td>
<td>399</td>
<td>1.50%</td>
<td>14</td>
<td>3.20%</td>
</tr>
<tr>
<td>Public Health</td>
<td>435</td>
<td>1.70%</td>
<td>13</td>
<td>3.00%</td>
</tr>
<tr>
<td>Missing</td>
<td>-</td>
<td>-</td>
<td>1</td>
<td>0.20%</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>26359</strong></td>
<td><strong>100%</strong></td>
<td><strong>435</strong></td>
<td><strong>100%</strong></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Status</th>
<th>Frequency</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Freshman</td>
<td>35</td>
<td>8.00%</td>
</tr>
<tr>
<td>Sophomore</td>
<td>49</td>
<td>11.30%</td>
</tr>
<tr>
<td>Pre-Junior</td>
<td>31</td>
<td>7.10%</td>
</tr>
<tr>
<td>Junior</td>
<td>60</td>
<td>13.80%</td>
</tr>
<tr>
<td>Senior</td>
<td>109</td>
<td>25.10%</td>
</tr>
<tr>
<td>Graduate /Master</td>
<td>101</td>
<td>23.20%</td>
</tr>
<tr>
<td>Doctorate</td>
<td>50</td>
<td>11.50%</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>435</strong></td>
<td><strong>100.00%</strong></td>
</tr>
</tbody>
</table>
Analysis

- Factor analysis
  - Reduces the number of variables
  - Examines correlations among observed variables
  - Identifies groups of interrelated variables (→ factors)
Factor Analysis in SPSS

- Data screening
- Analysis > Dimension reduction > Factor...
- Analysis settings
  - Variables: choose variables
  - Descriptive: KMO and Bartlett’s Test
  - Extraction: Principal component
  - Rotation: Varimax
  - Options
    - Sorted by size
    - Suppress small coefficients (below 0.35)
Initial 9 Factors

- Exploratory analysis
  - Some of 9 factors were non-intuitive

- How many factors?
  - Maximum variance explained by fewer factors
    - Total variance explained – Cumulative %
    - Number of factors
    - Useful and meaningful groups of variables

<table>
<thead>
<tr>
<th>Factor</th>
<th>Initial Eigenvalues</th>
<th>Extraction Sum of Squared Loadings</th>
<th>Rotation Sum of Squared Loadings</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Total % of Variance</td>
<td>Cumulative %</td>
<td>Total % of Variance</td>
</tr>
<tr>
<td>1</td>
<td>8.748</td>
<td>29.159</td>
<td>29.159</td>
</tr>
<tr>
<td>2</td>
<td>2.489</td>
<td>8.395</td>
<td>37.454</td>
</tr>
<tr>
<td>3</td>
<td>1.915</td>
<td>6.385</td>
<td>43.839</td>
</tr>
<tr>
<td>4</td>
<td>1.660</td>
<td>5.333</td>
<td>49.372</td>
</tr>
<tr>
<td>5</td>
<td>1.567</td>
<td>5.024</td>
<td>54.396</td>
</tr>
<tr>
<td>6</td>
<td>1.331</td>
<td>4.438</td>
<td>58.833</td>
</tr>
<tr>
<td>7</td>
<td>1.237</td>
<td>4.124</td>
<td>62.968</td>
</tr>
<tr>
<td>8</td>
<td>1.148</td>
<td>3.827</td>
<td>66.785</td>
</tr>
<tr>
<td>9</td>
<td>1.086</td>
<td>3.621</td>
<td>70.406</td>
</tr>
<tr>
<td>10</td>
<td>0.934</td>
<td>3.113</td>
<td>73.519</td>
</tr>
<tr>
<td>11</td>
<td>0.862</td>
<td>2.873</td>
<td>76.392</td>
</tr>
</tbody>
</table>
Final 5 Factors

- Total variance explained: 54.4%

- Meaningful components
  1. Printing, scanning, computers, and space
  2. Liaison’s help, events participation, staff help, and borrowing leisure books
  3. Circulation desk, textbooks on reserve, staff help, computers
  4. Online databases, ILL, library catalog, textbooks on reserve
  5. Library space use for event, group meeting, and study
Discovering User Attributes

- Save factor scores as variables
- Categorize cases based on the scores
- Crosstabs between the categorized cases and user characteristics (college, status, and location)
Persona Attributes

**Printing, Scanning & Computer use**
- Freshman
- Arts & Sciences
- Main Campus Library

**Liaison’s help via web**
- Senior
- Education
- Online

**Borrow reserve books in circulation desk**
- Doctorate
- Medicine
- Health Sci. Libraries

**Online databases and ILL**
- Master’s
- Biomedical Science
- Online

**Space use for group meeting and study**
- Sophomore
- Media Arts and Design
- Main Campus Library

* Images are from Drexel Libraries, Getty Images, and Google Image DB.
Limitations

1. Data might not support factor analysis
2. Interval and ratio scaled variables required
3. Larger sample size than qualitative research
4. Applicability of factor analysis limited by scope of survey questions
5. Experience and knowledge of your library services and users
Strategies & Recommendations

- Data-driven development
- Quantitative evidence mixed with qualitative methods
  - Apply to your library survey data
  - Validate previously developed personas
- Further analyses - cluster, crosstabs, ANOVA
- Interpret with more data
  - Circulation
  - Entrance
  - e-Resource usage
Thank You!
June 18 – 21, 2017

9th Evidence Based Library and Information Practice Conference

Eblip9.org