**Profiling the Different Types of Data Scientists: Which One is Right for You?**

**Sanaz Bahargam**

**Boston University**

**Theodoros Lappas**

**Stevens Institute of Technology**

---

**Motivation**

"The sexiest job of the 21st century"

- A hiring priority across industries.
- Ever-increasing demand, highly competitive salaries.
- What is it that makes a good data scientist?

**On the downside...**

- An emerging career with fuzzy requirements.
- Amazingly diverse, DSs with radically different backgrounds.
- Recruiting becomes a nightmare:

  - Which is the right DS type for your firm?

---

**The Skill Matrix**

<table>
<thead>
<tr>
<th>BIO</th>
<th>BSN</th>
<th>CS</th>
<th>ENG</th>
<th>HUM</th>
<th>MPS</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>BIO</strong></td>
<td>Bioinformatics, Molecular Biology, Statistics, Computational Biology, Genomics</td>
<td>Bioinformatics, Molecular Biology, Data Analysis, Statistics, R</td>
<td>Bioinformatics, Molecular Biology, Data Analysis, Statistics, R</td>
<td>Bioinformatics, Molecular Biology, Data Analysis, Statistics, R</td>
<td>Bioinformatics, Molecular Biology, Data Analysis, Statistics, R</td>
</tr>
<tr>
<td><strong>BSN</strong></td>
<td>Business Intelligence, SQL, Business Analysis, Project Management, Strategy</td>
<td>Data Analysis, Microsoft Excel, Analysis, SAS, Microsoft Office</td>
<td>SQL, Business Intelligence, Analysis, Microsoft Excel, Business Analysis</td>
<td>Business Analysis, SQL, Business Intelligence, Java, C++</td>
<td>Business Intelligence, Business Analysis, Database, Project Management, Management</td>
</tr>
<tr>
<td><strong>HUM</strong></td>
<td>SPSS, Research Design, Qualitative Research, Quantitative Research, Survey Design</td>
<td>Statistics, Research Design, Qualitative Research, Quantitative Research, Experimental Design</td>
<td>Statistics, Data Analysis, SPSS, R, Quantitative Research</td>
<td>SPSS, Statistics, Data Analysis, Qualitative Research, Quantitative Research</td>
<td>SPSS, Research Design, Qualitative Research, Quantitative Research, Experimental Design</td>
</tr>
</tbody>
</table>

---

**Methodology**

- Collect the full LinkedIn Profiles of 15K Data Scientists.
- We identify 6 large clusters with respect to undergraduate education: CS, Business, Bio, Math, Engineering, Humanities.

We study their differences in terms of:

1. **Skillsets as Data Scientists** (Pairwise Fisher Test)
2. **Occupational Commitment** (Survival Analysis)

- Define death as the transition to a non-DS career.
- Identity Features that are correlated with time-to-death.
- Project workers on the selected feature space.
- Use a clustering algorithm to identify sub-groups within each education-based cluster (DBSCAN).
- Plot the survival curve of each sub-group.

---

**Findings**

DSs with different educational backgrounds have significant differences with respect to:

- Their skillsets (Anticipated...)
- Their Occupational Commitment (Interesting!)

- Some types are more cohesive than others wrt their Occupational commitment.
- Clusters within types can be explained by their skills.
- Graduate Education increases commitment!