Analytics at the scale of amazon Prime
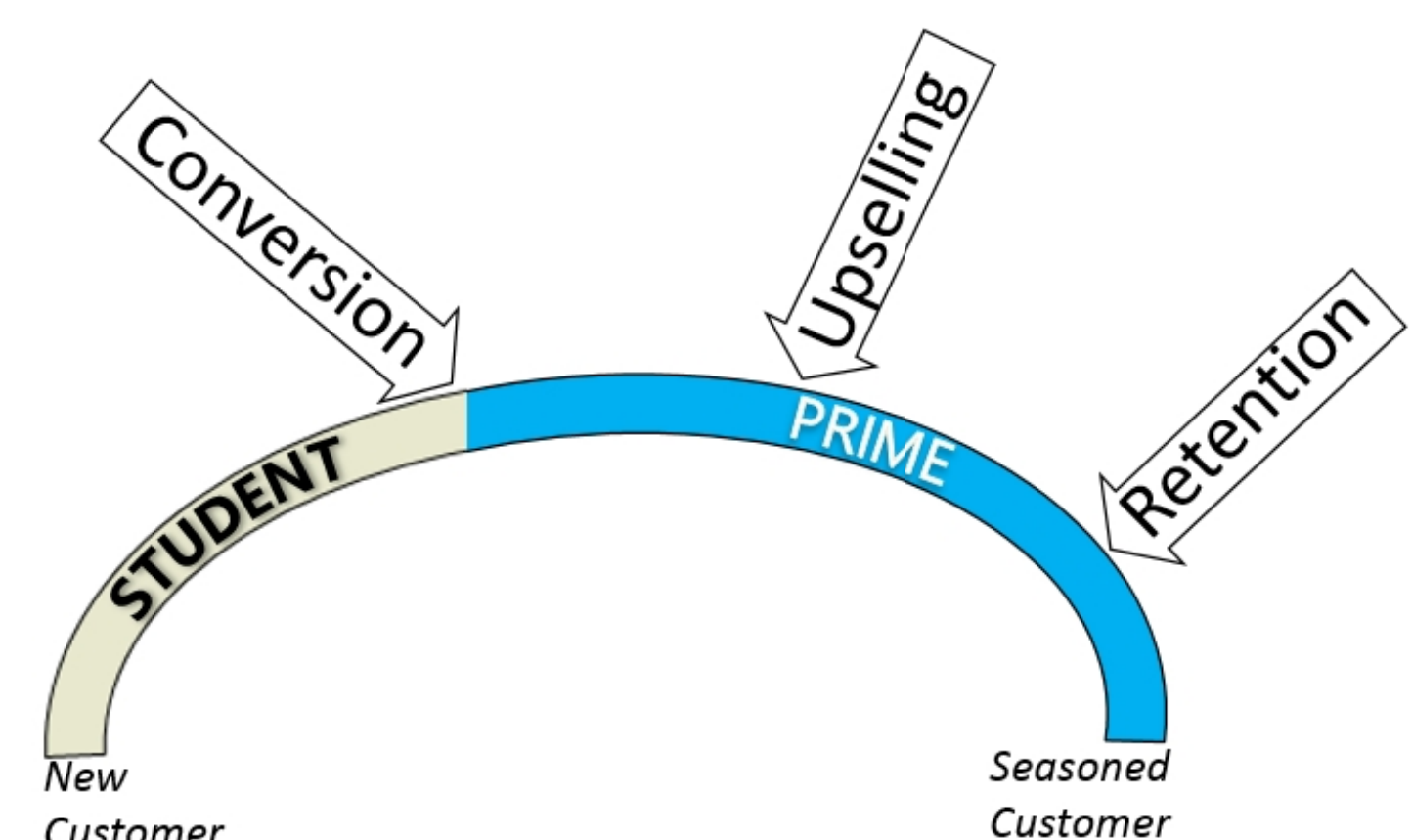
"Finding the needle in the haystack you didn't know you were looking for"

1. Our Goal

Find new Prime customer insights that add value to Amazon and the user

- Operations
- Marketing
- Product lines

2. Framework:

Drivers of the Prime Customer lifecycle

4. Taming the Beast

Feature engineering makes:
- Raw data meaningful
- Relevant to business question
- Manageable

Examples:
- Customer activity
- Product diversity & density
- Spending segments

3. The Data:

3 customer classifications reflecting tenure

~75 product groups

~9M records consisting of weekly observations for 170k customers

| Customer ID	Group	Week	Revenue, Product Group 1	Units, Product Group 1	Revenue, Product Group 2	Units, Product Group 2	Units, Product Group 75
1111	Student	6-Jan-13	0	0	5.75	2	41.40
1112	Student	6-Jan-13	0	0	0	0	0
1113	Older	6-Jan-13	14.12	3	0	0	11.87
1114	Newer	6-Jan-13	0	0	0	0	0
1111	Student	13-Jan-13	0	0	2.74	1	0
1112	Student	13-Jan-13	0	0	0	0	0
1113	Older	13-Jan-13	30.24	8	5.23	1	0
1114	Newer	13-Jan-13	0	0	0	0	3.3
1111	Student	19-Jan-13	6.9401	2	0	0	0
1112	Student	19-Jan-13	0	0	0	0	1.10
...	...	...	...	...	...	...	...

5. Insights:

A. Data exploration (i.e. find leads)
   A. Visualize data
   B. Characterize features
   C. Differentiate further
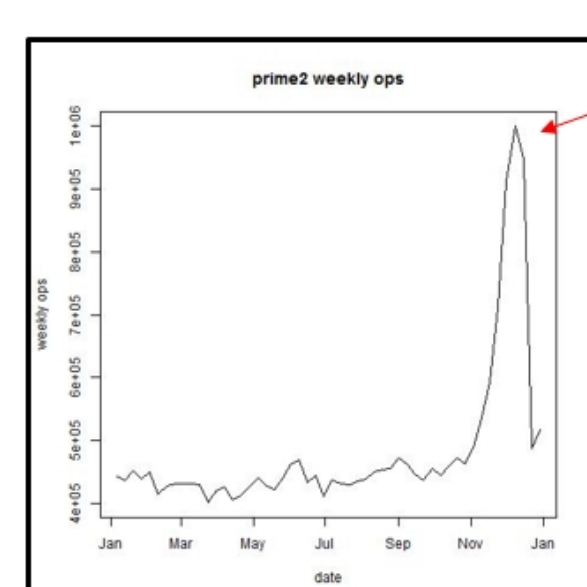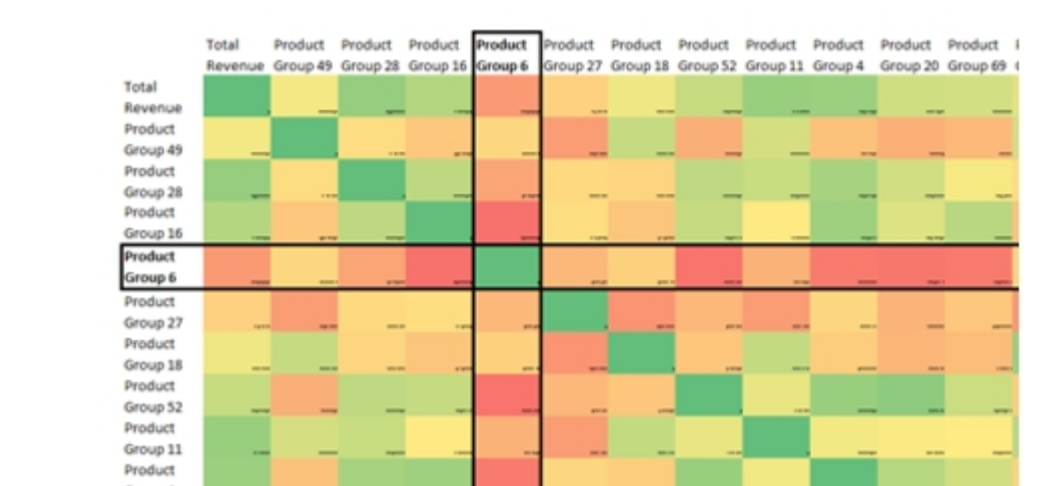

Prime Customer Segment
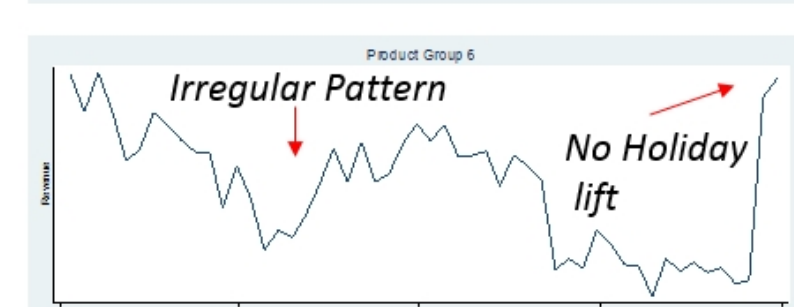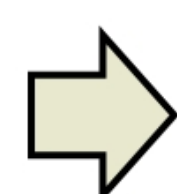
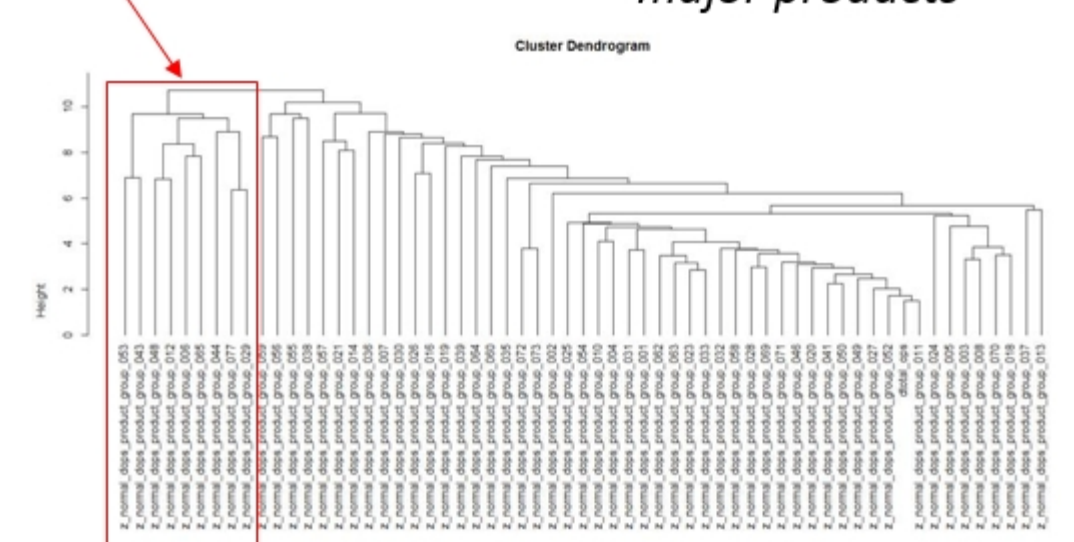Prime Customer Segment differentiated as Product Groups

B. Validate differentiators
   - Correlation tables
   - Clustering
   - Two-way Tables
   - Regressions
   - Decision trees

C. Quantify effects and make inferences about how these relate to Prime lifecycle drivers

Product Group 6 in isolated branch

Product Group 6 has low / negative correlation w/ all major products

Key differentiators between customer segments

Team: Brandon Bartch, Zach Freeman, Jarek Langer, Thomas Stephens
Advisors: George Westerman, Sagit Bar-Gill, Shachar Reichman