Predicting Student Performance in edX MOOCs using Browser Events

By Joshua E. Jodesty

Objectives & Acquisition:

- Predict student outcomes in MOOCs offered by edX
  - Outcomes: Certification & Grade Ranges
- Determine browser events that affect student outcomes
- Dataset: 129843 student logs containing browser events
- Source: edX courses administered by HarvardX & MITx from 2012 & 2013

Methodology:

- Cleansing
- Exploration: Generated 15 New Features
- Modeling & Evaluation
- Repeat

Target Course

Title: Intro. to Computer Science & Programming

Institution: MITx

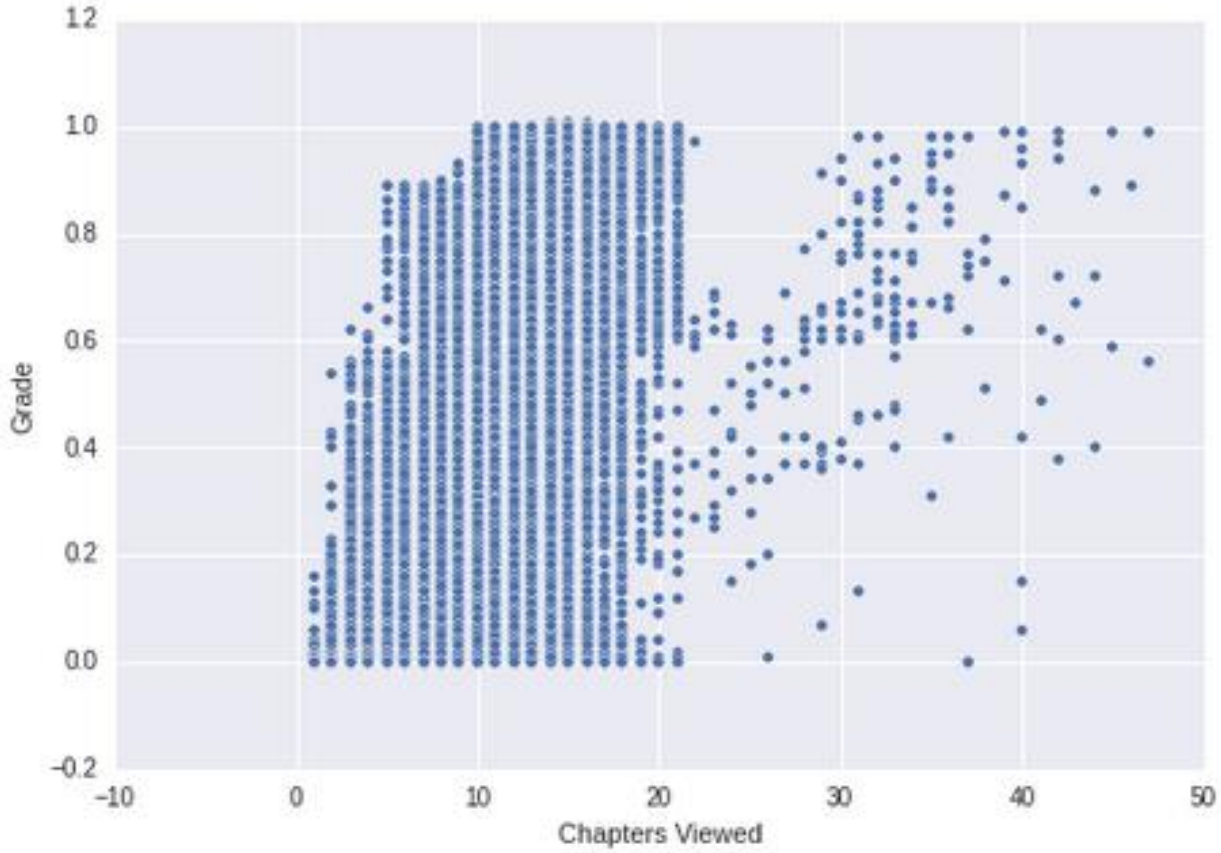
Number: 6.00x

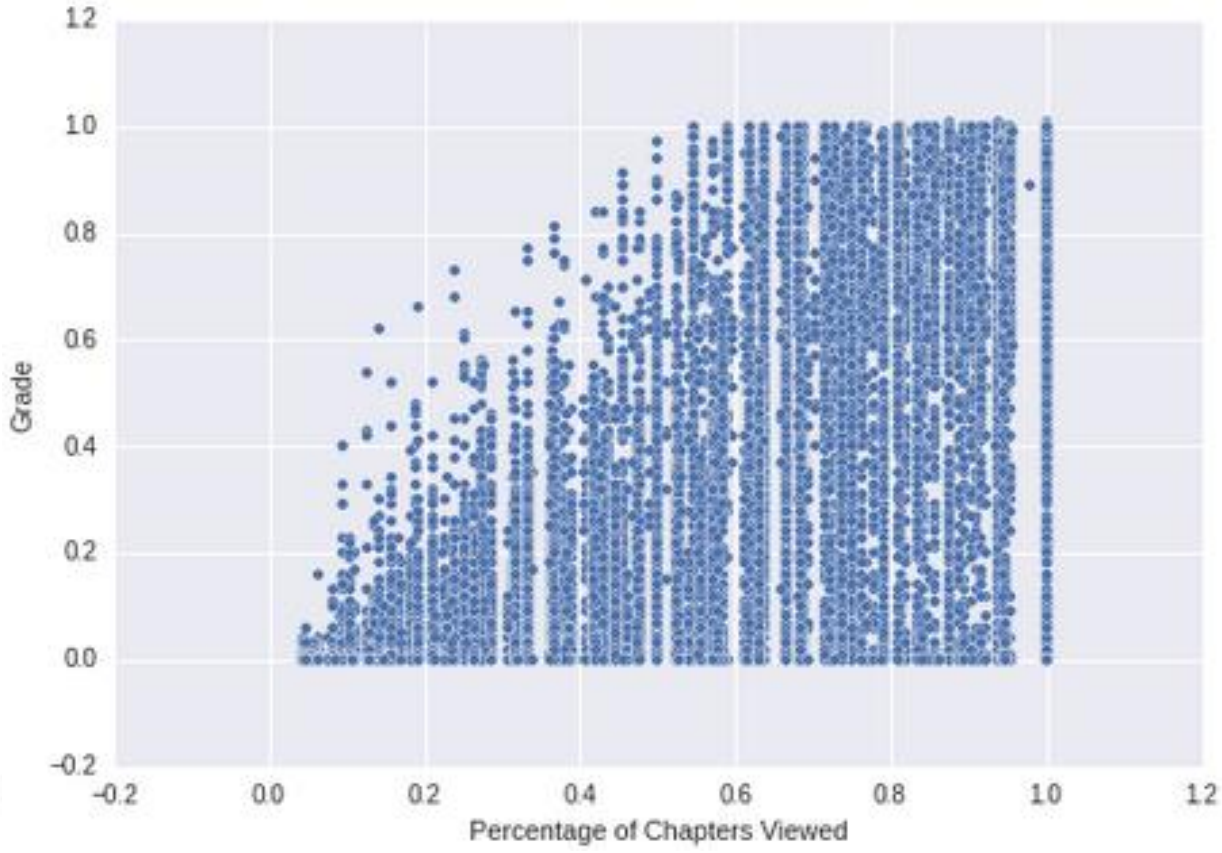
Logs: 51437

Certified: 5 %

2539 Certifications

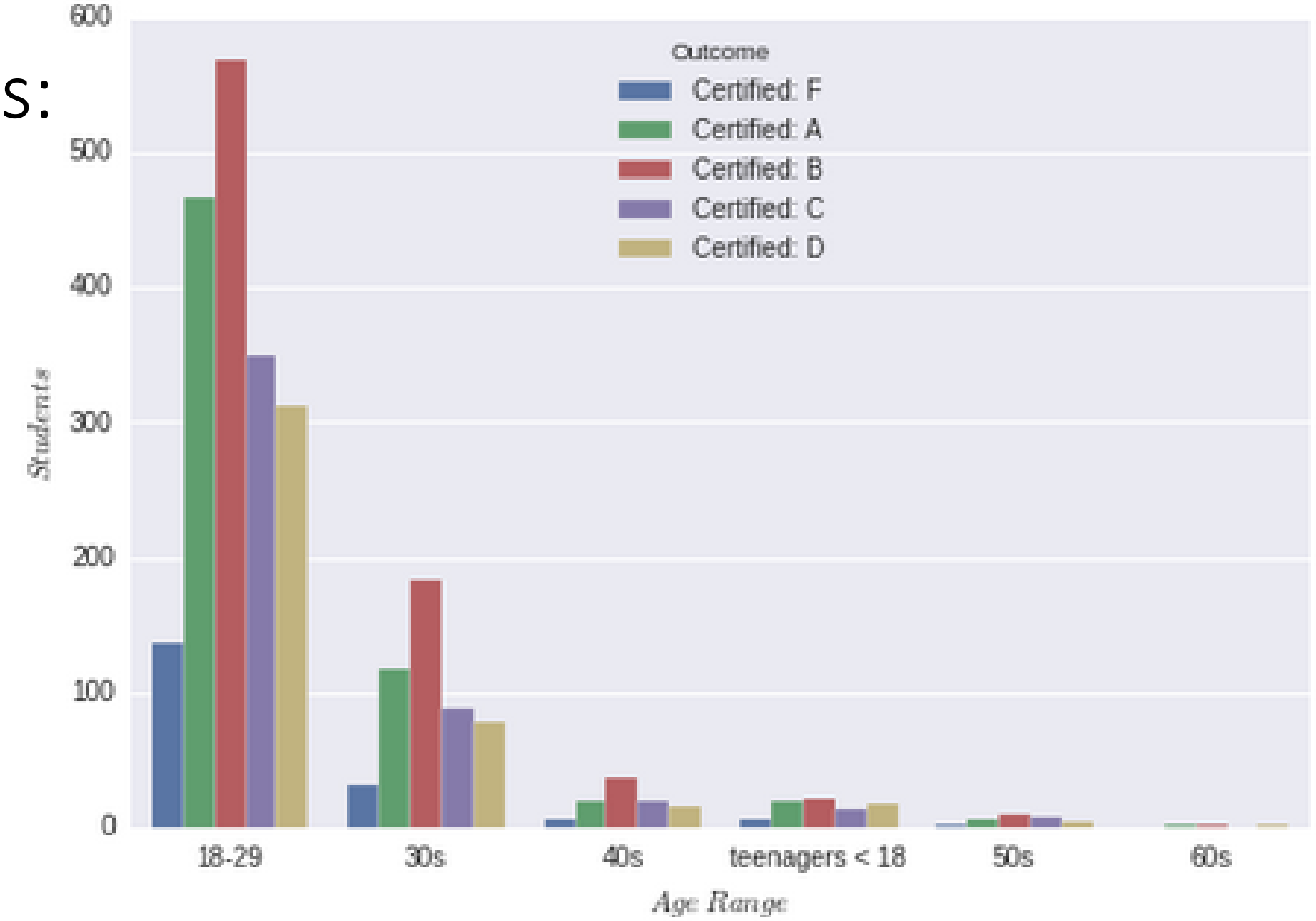
Discoveries: Tranformation



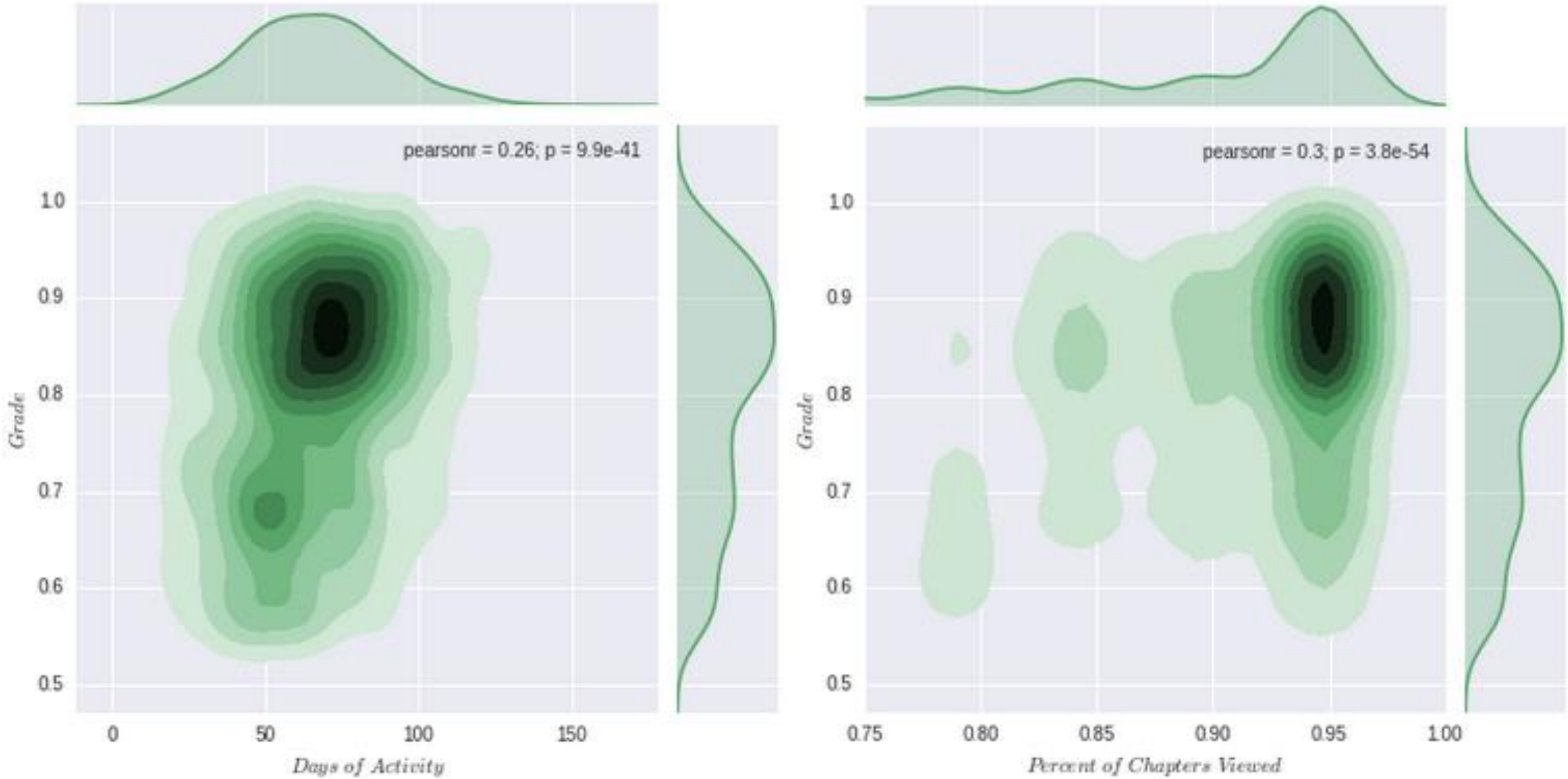


Discoveries:

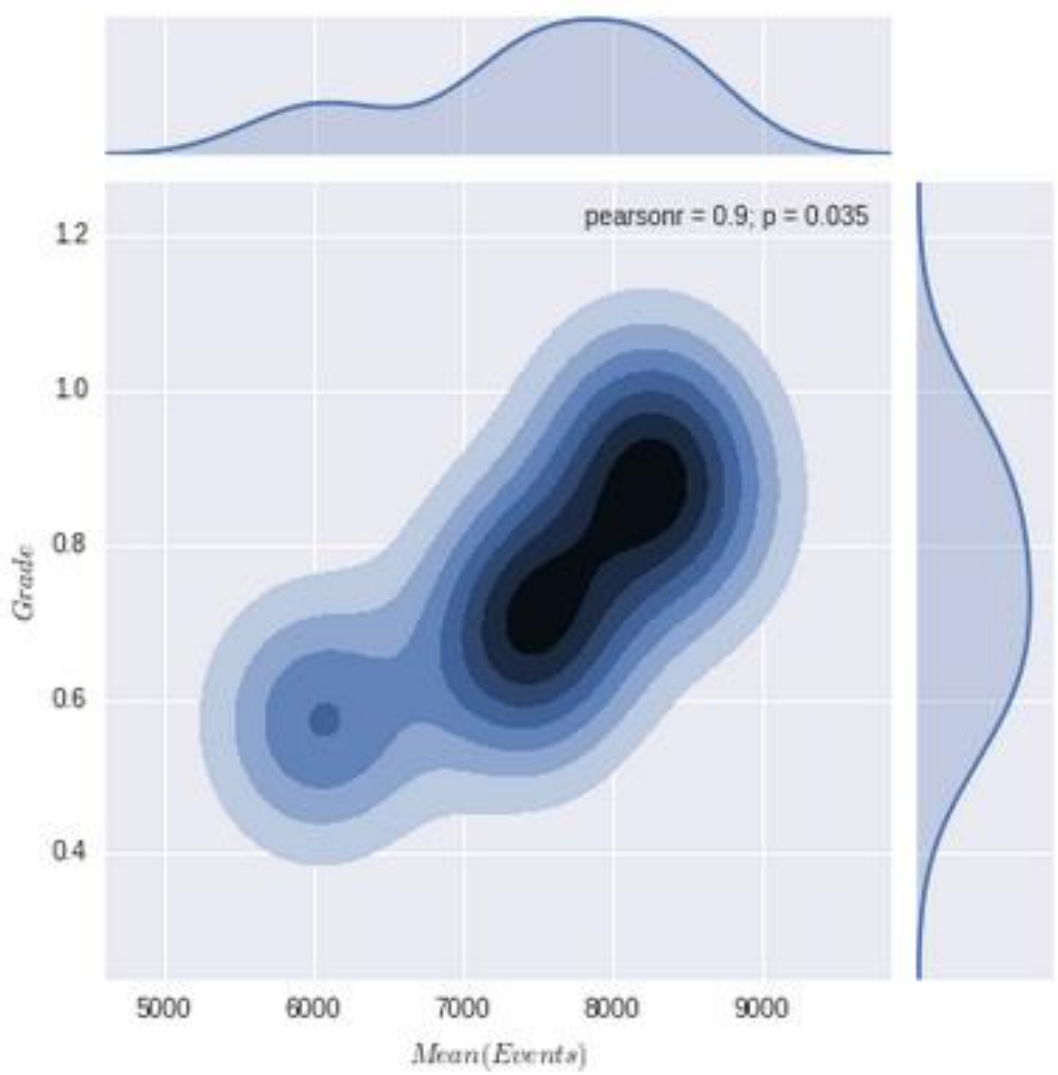
- Student Outcome count by Age Range

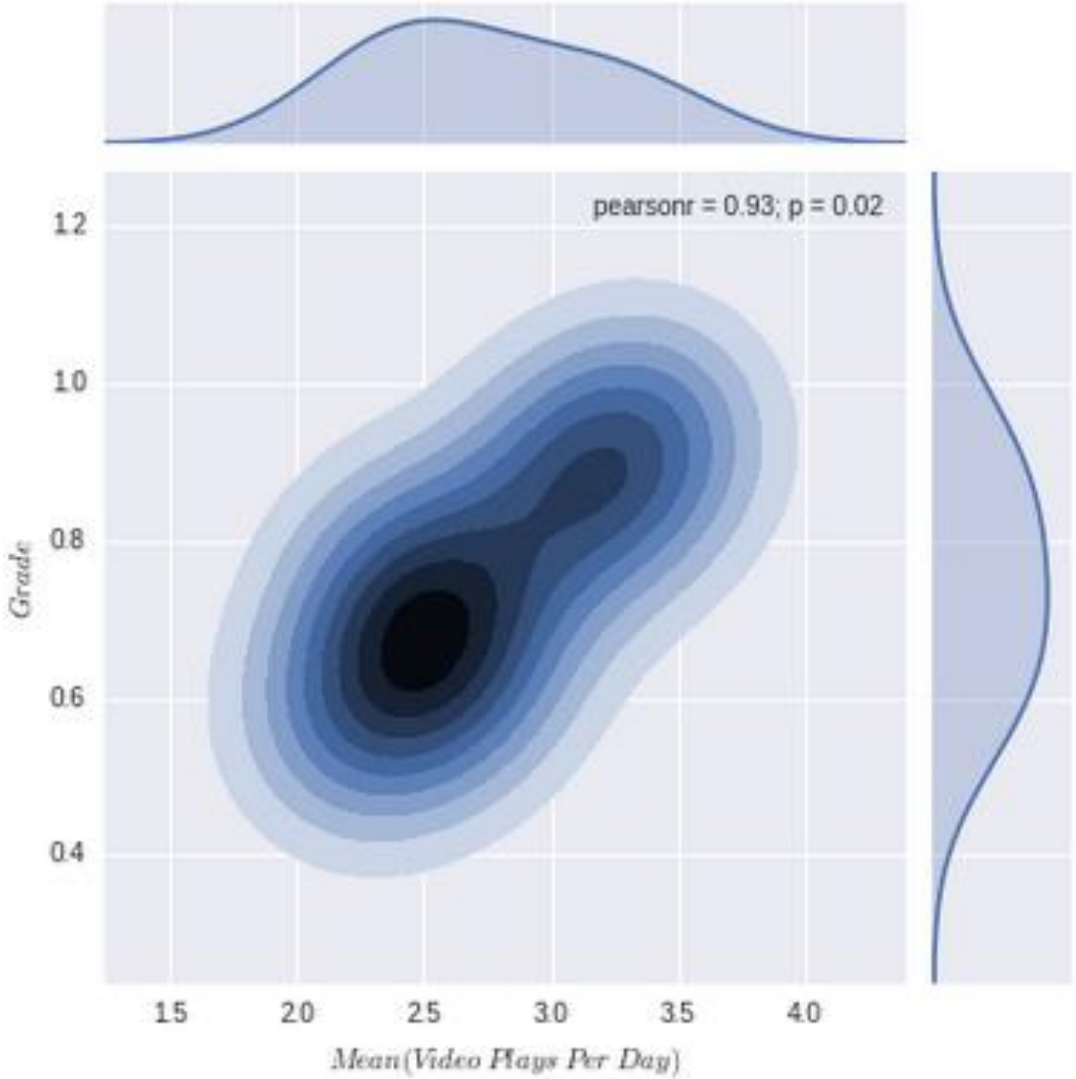


Discoveries: Student Logs



Discoveries: Student Outcomes

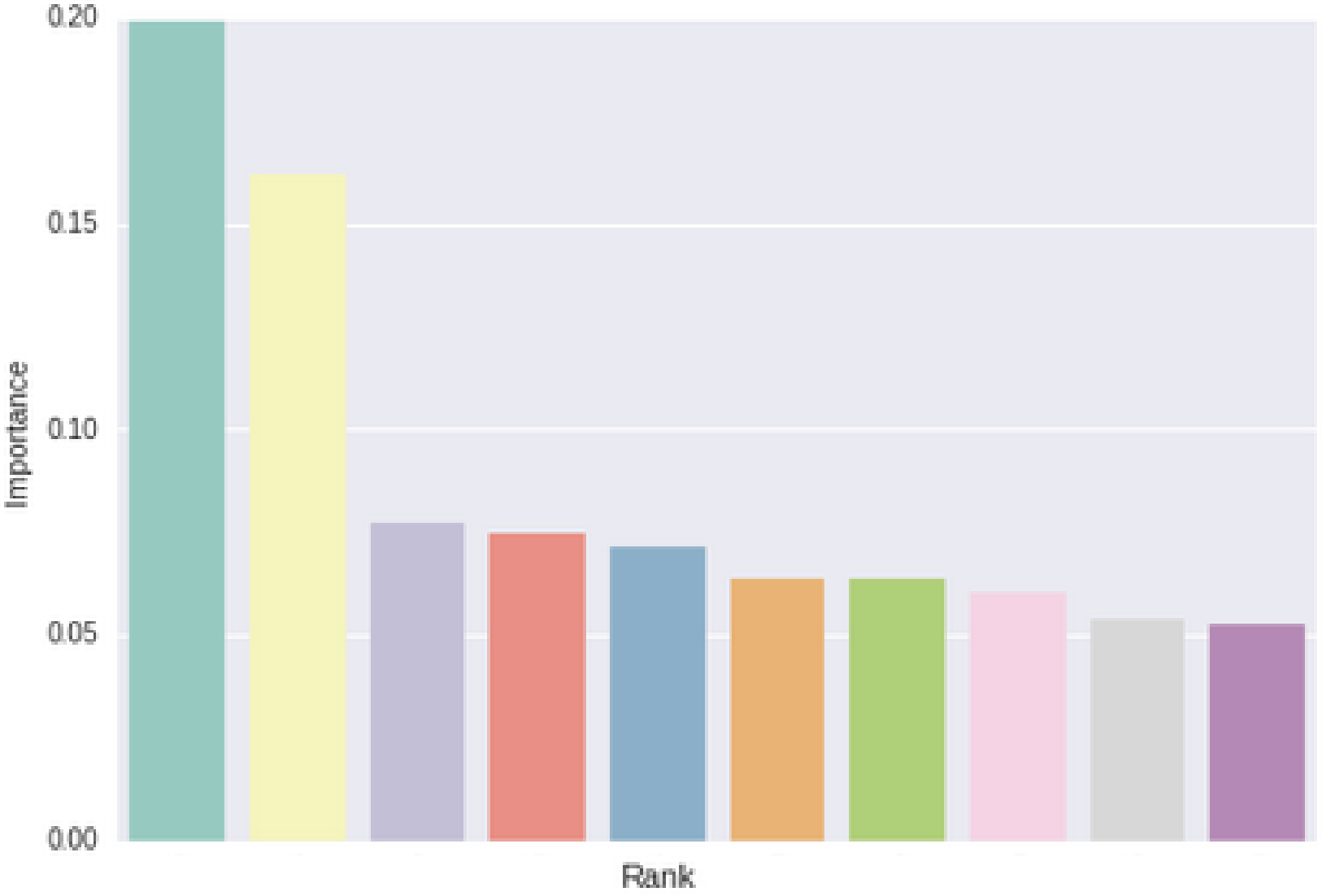




Modeling: Random Forest Classifier

- Train: MITx dataset without 6.00x
- Test: 6.00x
- Most Relevant Metrics:
  - Accuracy Score: 0.995
    - 95.6% of students correctly classified (49149)
  - Zero One Loss: 0.005
    - 4.5% of students misclassified (2288)

Modeling: Top 10 Important Features



Feature Importance Rank:

1. Percentage of Chapters Viewed
2. Chapters Views per Day
3. Browser Events per Day
4. Days of Video Activity
5. Days of Activity
6. Video Plays / Events
7. Events
8. Chapter Views
9. Video Plays
10. Video Plays per Day

Questions?

Contact:

linkedin.com/in/joshuajodesty

@jejodesty

github.com/JEJodesty