Talent Analytics

Kaz Hayashi
Tony Xu
Chien Chen Liew
Arnav Gupta

Agenda

Part 1

Problem Statement and challenge

Data preprocessing

Modelling

Results

Inference (Threats to validity)

Improvements and Next steps

Part 2

Current landscape of some tools

Potential Concerns

Our proposed Tool

Part 1

- What are the organizational and social factors associated with examiner attrition?
- What is the role of gender, race and ethnicity here in the processes underlying the question above?

Data Preprocessing

- + Gender, Race, Tenure days
- Removed gender NA values
- + Application quantity related variables (new, pending, issued, etc.)
- + Quarter variable from application date
- + Art unit mobility & separation variables
- Aggregated by examiner_id and quarter (Panel Data)

- Aggregate the data by examiner_id
- + Variables like “ration of woman, minority, their own race” in their unit

Why we don't use panel data structure

– because some examiner moved art units multiple times in a quarter.

	examiner_id	quarter	examiner_art_unit
72673	59012.0	2006/3	1716.00
72690	59012.0	2006/3	1717.00
73589	59012.0	2006/3	1792.00
76494	59012.0	2006/4	1716.00
76511	59012.0	2006/4	1717.00

- This caused the dataset to have multiple observations for the same examiner_id and quarter, which is not ideal for panel data structure.

Why we don't use panel data structure

In Fixed Effects model:

$$Y_{it} = \alpha_i + \beta X_{it} + \epsilon_{it}$$

Where:

- Y_{it} is the dependent variable for individual i at time t,
- α_i represents the individual-specific effect (fixed effect) that captures all time-invariant characteristics of the individual,
- X_{it} is a vector of time-varying explanatory variables with coefficients β,
- ϵ_{it} is the idiosyncratic error term.

So, time-invariant variables' effects like "Gender" and "Race" will be absorbed by α

Overview of Data

Selected Variable Name	Explanation
start_year	The year the examiner started working at the USPTO.
Separation_indicator (Y)	Sum indicator of whether the examiner has left the USPTO (e.g., 1 for yes, 0 for no).
AU_move_indicator (mobility)	Sum indicator of the number of times the examiner has changed art units.
Avg_woman_ratio	Average ratio of women (career wise)
Avg_minority_ratio	Average ratio of minority individuals in the art unit (career wise)
Own_race_ratio	Ratio of examiners within the art unit sharing the same racial or ethnic background as the examiner. (career wise)

Modelling - Logistic Regression

Model 1: Basic Demographics

$$\log\left(\frac{P(Y=1)}{1-P(Y=1)}\right) = \beta_0 + \beta_{\text{gender}} + \beta_{\text{race}} + \beta_{\text{tenure}} + \beta_{\text{start}}$$

Model 2: Including Application Metrics

$$\log\left(\frac{P(Y=1)}{1-P(Y=1)}\right) = \text{Model 1} + \beta_{\text{new\_apps}} + \beta_{\text{issued\_apps}} + \beta_{\text{abn\_apps}} + \beta_{\text{pen\_apps}}$$

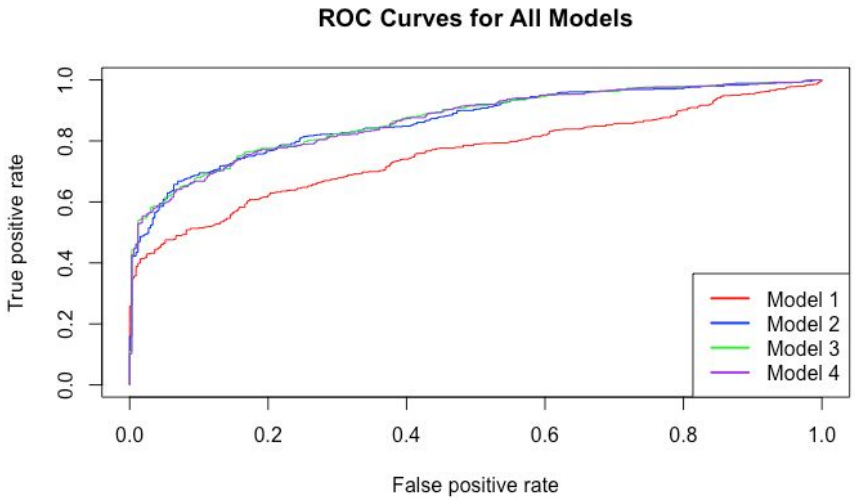
Model 3: Adding Art Unit Information

$$\log\left(\frac{P(Y=1)}{1-P(Y=1)}\right) = \text{Model 2} + \beta_{\text{au\_move}} + \beta_{\text{avg\_unit}} + \beta_{\text{woman\_ratio}} + \beta_{\text{minority\_ratio}} + \beta_{\text{own\_race}}$$

Model 4: Including Interactions

$$\log\left(\frac{P(Y=1)}{1-P(Y=1)}\right) = \text{Model 3} + \beta_{\text{gender}\times\text{woman\_ratio}} + \beta_{\text{gender}\times\text{minority\_ratio}}$$

Results and Discussion



Model	**AUC**
Model 1	0.75570256556172
Model 2	0.75570256556172
Model 3	0.868549343197229
Model 4	0.866861099255464

Comparative Logistic Regression Model Summary

	Dependent variable:			
	separation_indicator_sum			
	(1)	(2)	(3)	(4)
gendermale	0.130	0.096	0.052	0.637
	(0.080)	(0.086)	(0.096)	(0.403)
raceblack	-0.100	-0.012	-0.114	-0.133
	(0.215)	(0.228)	(0.245)	(0.245)
raceHispanic	0.149	0.180	0.053	0.044
	(0.203)	(0.216)	(0.236)	(0.236)
raceother	0.068	0.084	-0.022	0.002
	(1.431)	(1.460)	(1.516)	(1.517)
racewhite	0.067	0.176*	0.123	0.121
	(0.086)	(0.093)	(0.140)	(0.140)
tenure_days	-0.003***	-0.002***	-0.001***	-0.001***
	(0.0002)	(0.0002)	(0.0002)	(0.0002)
start_year2001	-1.665***	-1.116***	-0.982***	-0.986***
	(0.178)	(0.188)	(0.187)	(0.187)
start_year2002	-2.929***	-1.881***	-1.554***	-1.558***
	(0.236)	(0.246)	(0.238)	(0.238)
start_year2003	-3.968***	-2.406***	-1.807***	-1.802***
	(0.298)	(0.314)	(0.293)	(0.293)
start_year2004	-5.145***	-2.987***	-2.287***	-2.287***
	(0.367)	(0.388)	(0.358)	(0.358)
start_year2005	-5.926***	-3.171***	-2.198***	-2.205***
	(0.442)	(0.469)	(0.429)	(0.430)
start_year2006	-6.944***	-3.481***	-2.276***	-2.269***
	(0.527)	(0.561)	(0.512)	(0.512)
start_year2007	-7.569***	-3.590***	-2.166***	-2.171***
	(0.603)	(0.648)	(0.594)	(0.594)
start_year2008	-8.633***	-3.912***	-2.202***	-2.188***
	(0.690)	(0.745)	(0.683)	(0.683)
start_year2009	-10.219***	-4.958***	-2.914***	-2.922***
	(0.774)	(0.828)	(0.757)	(0.758)
start_year2010	-11.010***	-5.285***	-2.968***	-2.960***
	(0.835)	(0.893)	(0.814)	(0.815)
start_year2011	-11.522***	-5.005***	-2.464***	-2.461***
	(0.906)	(0.982)	(0.897)	(0.898)
start_year2012	-12.897***	-5.648***	-2.876***	-2.875***
	(0.985)	(1.065)	(0.972)	(0.972)
start_year2013	-13.752***	-5.281***	-2.202**	-2.216**
	(1.085)	(1.195)	(1.102)	(1.103)
start_year2014	-12.827***	-2.262	1.223	1.223
	(1.327)	(1.508)	(1.437)	(1.437)
start_year2015	-14.473***	-2.438	1.524	1.542
	(1.394)	(1.849)	(1.894)	(1.904)
start_year2016	-4.362	5.200	10.219	10.264
	(378.372)	(183.784)	(303.356)	(303.376)
new_applications_mean		-0.749***	-0.815***	-0.816***
		(0.049)	(0.052)	(0.052)
ISSUED_applications_mean		0.666***	0.689***	0.690***
		(0.049)	(0.051)	(0.051)
abn_applications_mean		0.684***	0.642***	0.647***
		(0.067)	(0.069)	(0.069)
PEN_applications_mean				
au_move_indicator_sum			-0.024***	-0.024***
			(0.002)	(0.002)
avg_num_in_art_unit			0.028***	0.029***
			(0.006)	(0.007)
avg_woman_ratio			-0.598	-0.192
			(0.370)	(0.618)
avg_minority_ratio			-1.078***	-0.248
			(0.352)	(0.571)
own_race_ratio			-0.237	-0.226
			(0.314)	(0.314)
gendermale:avg_woman_ratio				-0.609
				(0.736)
gendermale:avg_minority_ratio				-1.137*
				(0.615)
Constant	18.955***	12.760***	9.455***	9.029***
	(1.341)	(1.394)	(1.278)	(1.311)
Observations	3,880	3,880	3,880	3,880
Log Likelihood	-2,074.028	-1,872.100	-1,776.118	-1,774.400
Akaike Inf. Crit.	4,194.057	3,796.200	3,614.237	3,614.800
Note:				*p<0.1; **p<0.05; ***p<0.01

Across the models, significant predictors are:

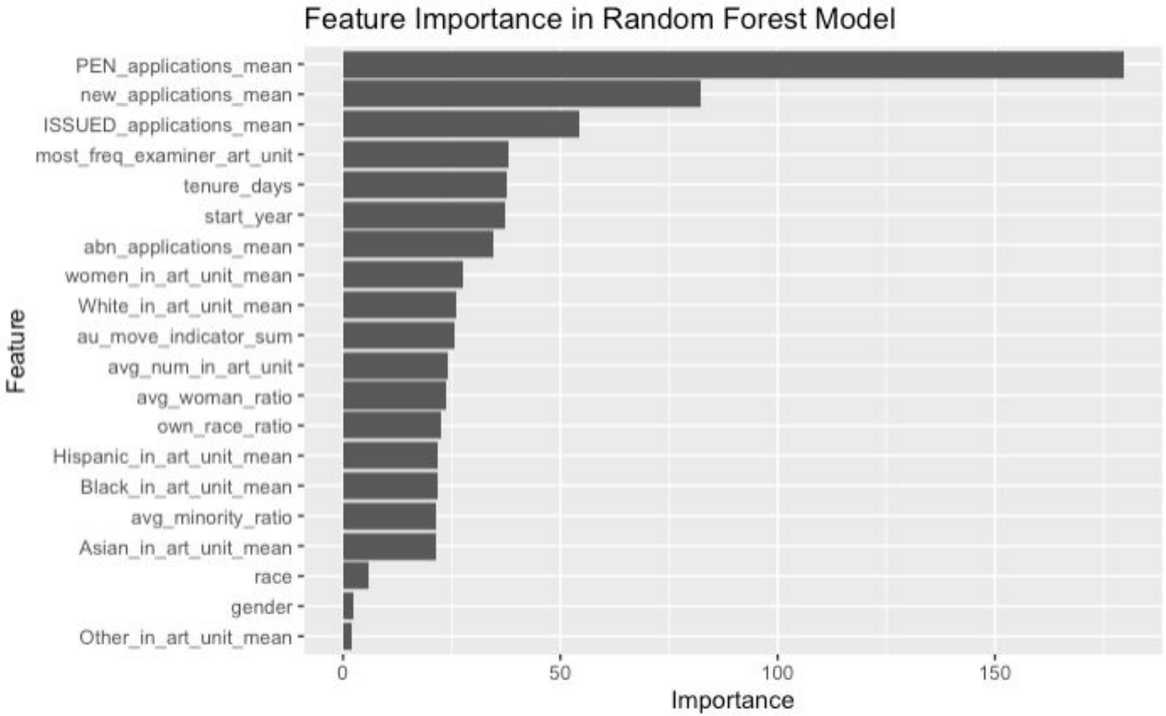
- Tenure days
- New_applications_mean (per quarter),
- issued_applications mean (per quarter)
- abn_applications mean (per quarter)

Across models 3 and 4, significant predictors are:

- Au_move_indicator_sum
- Avg_num_in_art_unit

Distinguishing between models 3 and 4, avg_minority_ratio is significant with a large magnitude.

Results and Discussion



Significant predictors validated with Random Forest Model's feature importance.

Conclusion and Remarks

- No clear relationship between gender, race/ ethnicity and examiner attrition. But it seems to suggest an increase in minority ratio reduces attrition.
- Significant predictors of attrition is their mean of applications (issued, abandoned, pending etc).
- Seems that average number of examiners in Art Unit plays a part in attrition. Could be possible that the environment in the different art unit affects examiner's desire to leave.

- However, the analysis could be biased by outliers through means to aggregate the data.
- Our model assumes a fixed effect model where other confounding variables might not be captured in the data. Requires a more dedicated experiment to determine the effects of Art Unit/ gender, race/ ethnicity.

Part 2

People Analytics Offerings Landscape

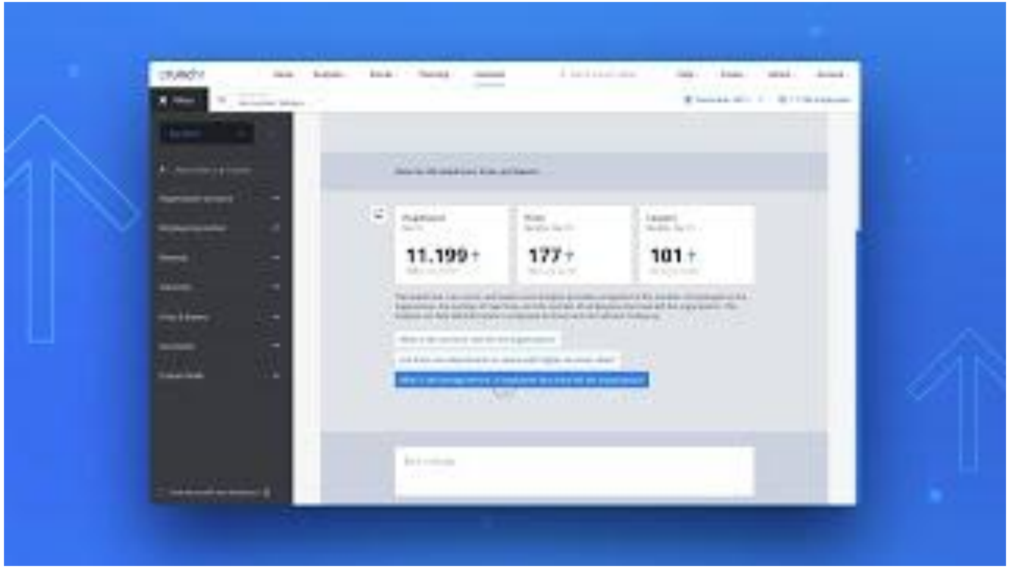
- There's quite a few people analytics offerings each with their own strengths

- One that stands out for our use case is **Crunchr**

Crunchr

- Uses AI to identify associations and causal relationships between the workforce and organizational objectives

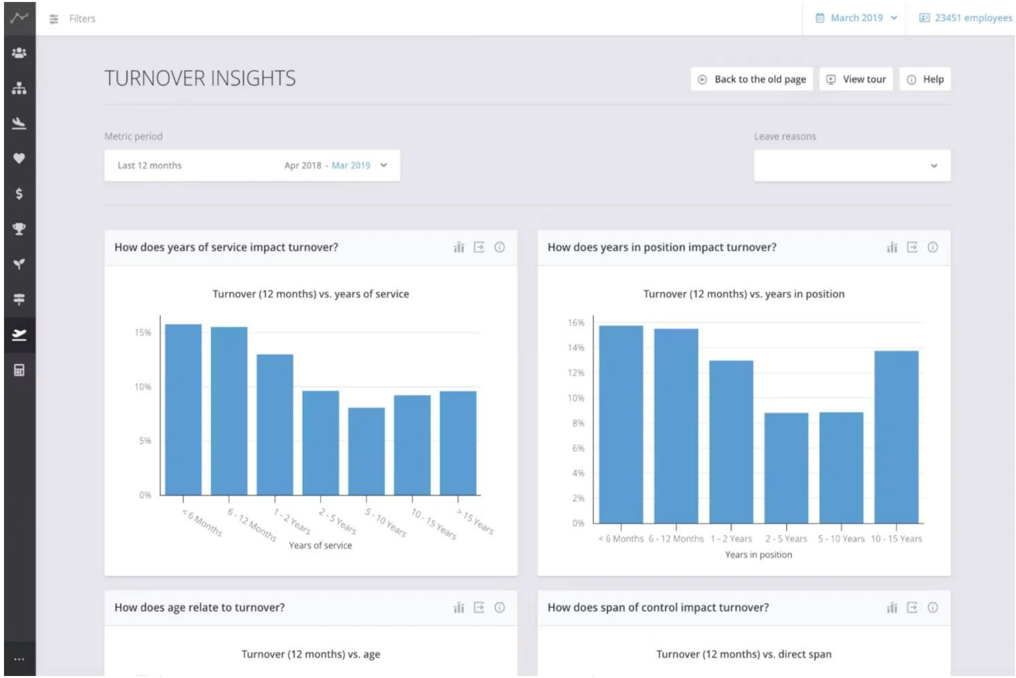
- Crunchr Assistant, an AI-powered digital co-pilot



https://www.youtube.com/watch?v=ndP7Og910TM&t=43s&ab_channel=Crunchr

Applicability to our use case

- **Segment and analyze** workforce data by demographics, performance, and relevant factors to identify organizational and social causes of attrition, such as workload, job satisfaction, engagement, and leadership styles.
- Utilize predictive analytics to **forecast attrition risks** based on historical data, enabling proactive interventions and
- Employ **benchmarking to compare attrition rates with industry standards**, aiding in determining whether trends are specific to the organization or reflective of broader industry patterns.



Potential Concerns

- **Data Quality and Bias:** The accuracy and impartiality of Crunchr's analytics hinge on data quality. Biased or incomplete historical data can perpetuate inequalities and inaccuracies in decision-making.
- **Algorithmic Transparency:** Concerns arise from the lack of transparency in how Crunchr's AI generates insights. Understanding the rationale behind analytics is crucial for justifying workforce decisions.
- **Over-reliance on Technology:** There's a risk of overlooking human judgment by solely relying on Crunchr for decision-making. Balancing technological insights with human expertise is essential, especially for organizations dealing with complex matters like the USPTO.
- **Interpretation and Actionability**: The ability to draw meaningful conclusions from the data depends on the context and understanding of the organizational culture. There's a risk of misinterpretation or oversimplification of complex social and organizational dynamics.
- **Privacy and Ethical Considerations:** Implementing Crunchr involves handling sensitive data, prompting privacy and ethical concerns. Complying with privacy laws and ethical standards is crucial for maintaining trust within the organization.

Our proposed Tool - PatentLens

The GenAI tool aims to solve macro to micro evaluations -

- **Volume Overload** - Cognitive overload from the sheer number of applications
- **Assessment Variability** - Inconsistent evaluations and depth over time

Implementation Pillars -

Ensures model remains dynamic and adaptable to current conditions

Guard Rail Conditioning

Establishes robust evaluation criteria tailored to patent examination, leveraging legal standards and technological benchmarks

Dynamic Content Archive

A curated, dynamically updated knowledge base, including recent patents, legal rulings, and technological advancements, to refine the AI's training and insights.

Active Trend Assessment

Integrates a search API to keep abreast of evolving market trends, technology landscapes, and legal precedents, informing the AI's analysis and recommenda-
tions.

Our proposed Tool - PatentLens

- Similar to our previous hackathon tool