Prescriptions and Manufacturer Payments

Zak Scholl – March 19, 2017

The Problem

- **Drug and device manufacturers pay healthcare providers**
- **Physicians received $2.6 billion in 2014**
- **Do these payments influence the rate at which drugs are prescribed?**
- **Who cares?**
  - US Department of Health and Human Services

The Data

- **Claims filed for Medicaid Part D**
  - Center for Medicaid and Medicare Services

	NPI	SPECIALTY_DESC	DRUG_NAME	TOTAL_DAY_SUPPLY
0	1821285826	Urology	TAMSULOSIN HCL	360
1	1093969024	Internal Medicine	PANTOPRAZOLE SODIUM	360
2	1518048750	Pediatric Medicine	VENLAFAXINE HCL ER	360
3	1952310666	Psychiatry	ABILIFY	420
4	1952310666	Psychiatry	ALENDRONATE SODIUM	480

- **Payments**
  - Open Payments

	Physician_Profile_ID	Total_Amount_of_Payment_USDollars	Applicable_Manufacturer_or_Applicable_GPO_Making_Payment_Name
0	269814	17.14	FOREST PHARMACEUTICALS, INC.
1	180504	11.27	FOREST PHARMACEUTICALS, INC.
2	180504	2.20	FOREST PHARMACEUTICALS, INC.
3	86548	47.70	FOREST PHARMACEUTICALS, INC.
4	307584	12.01	FOREST PHARMACEUTICALS, INC.

How do we solve?

- Can we predict rates of prescription *better* with payment data than without?
- Build two models
  - Reasonable predictive power
- Compare error rates

Data Wrangling

- **Have to combine both datasets**
  - No common identifier
- **Use NPPES**
  - National provider registry
- **Lookup identifiers with Name, Address**
  - Throw away any ambiguous matches
- **Left with 36% of claims data**
  - ~8.7 million rows

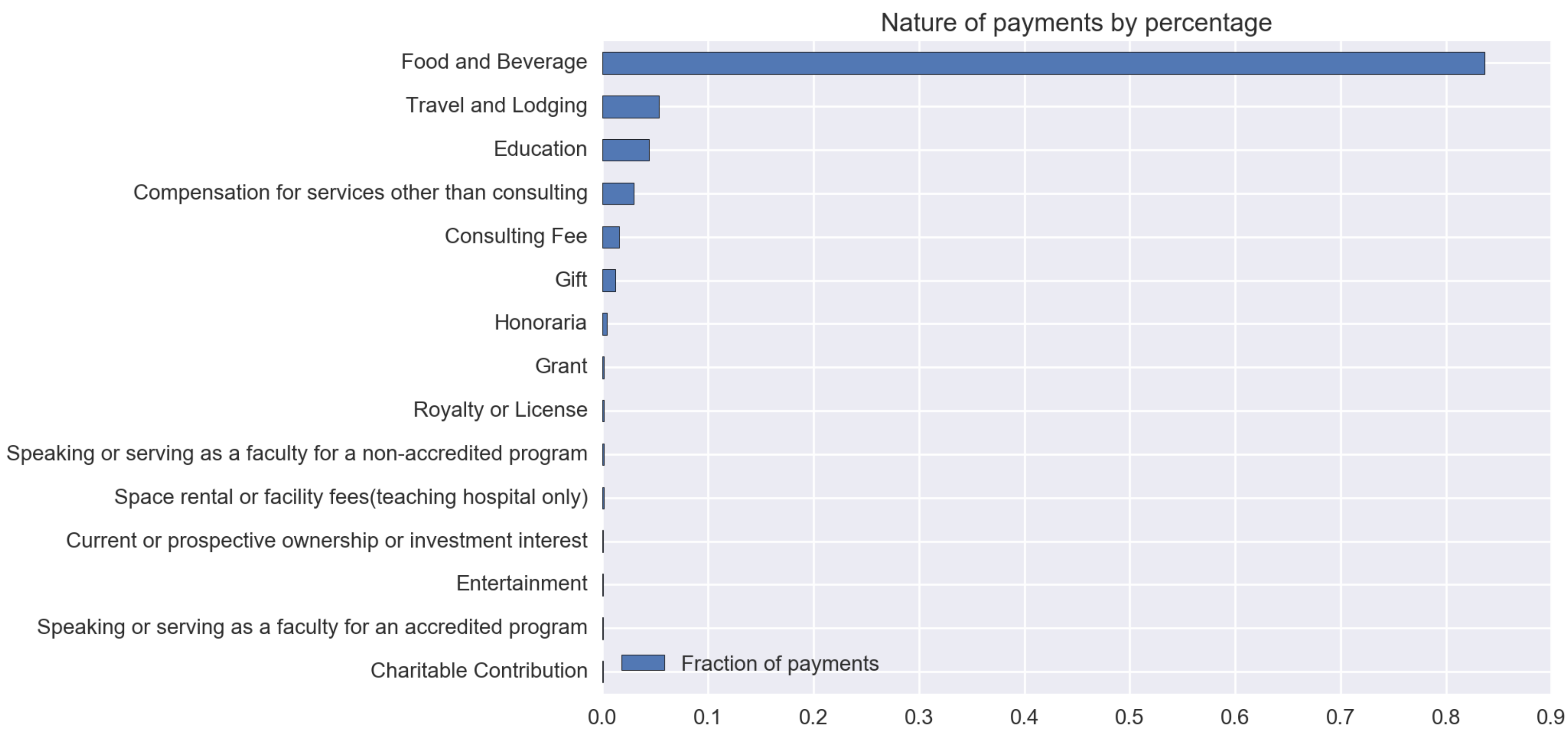
Data Wrangling

- **What to do with payments?**
- **Group by physician and manufacturer**
  - Sum total
- **Create a feature for every manufacturer**
- **Tie together with identifier**

ABIOMED	ABL Medical, LLC	...	iCAD, Inc	iRhythm Technologies, Inc.	iScreen Vision Inc.	integrated dental systems	nContact Surgical, Inc	optos plc	rEVO Biologics, Inc.	sanofi-aventis U.S. LLC	Physician_Profile_ID	NPI
0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100000	1215900089
0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1000001	1356429617
0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1000014	1366438970
0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100002	1215928759
0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100003	1215928916

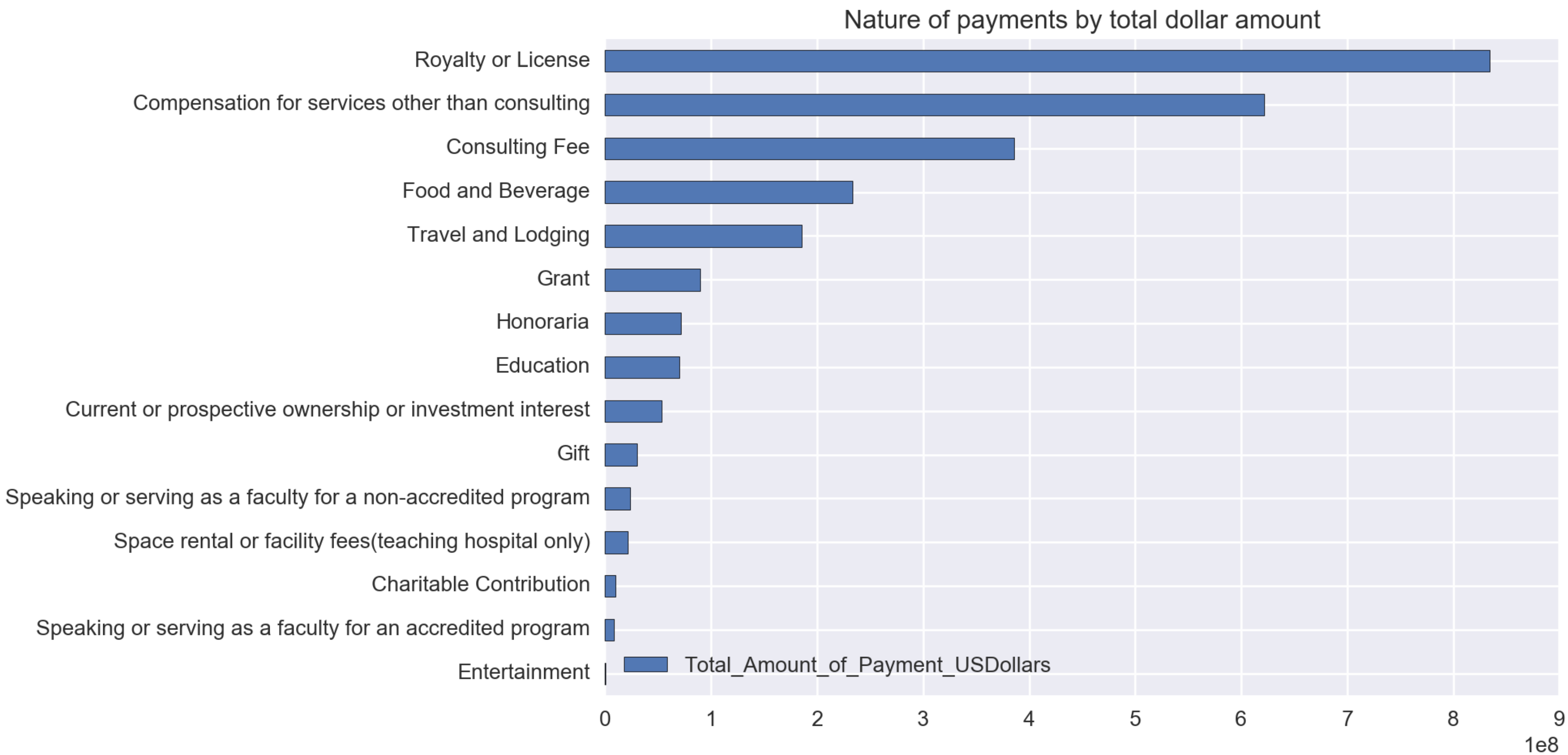
Exploration

- **Most payments are for Food and Beverage**



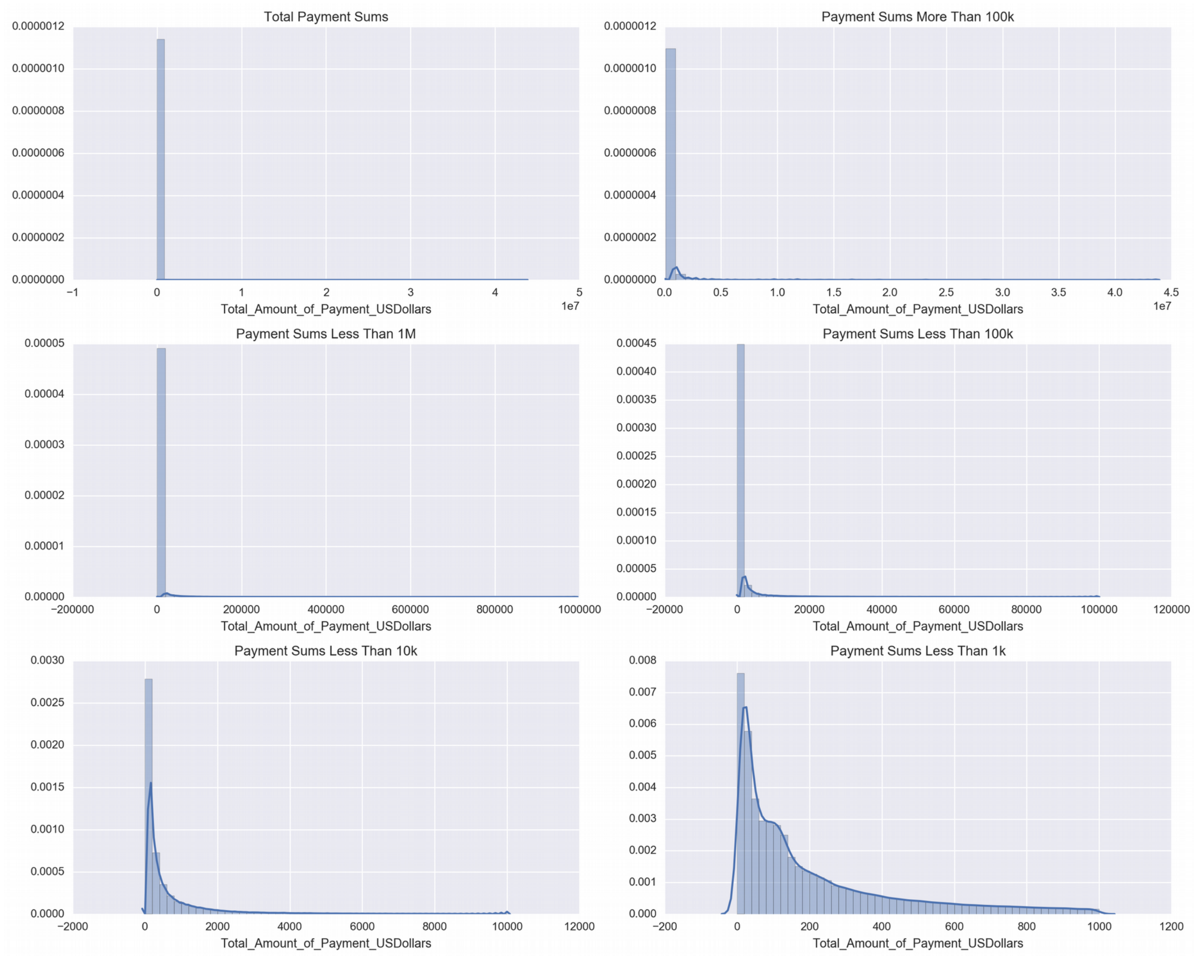
Exploration

- **Most $$$ is for consulting and royalties**



Exploration

- **Large variability in payment sums**

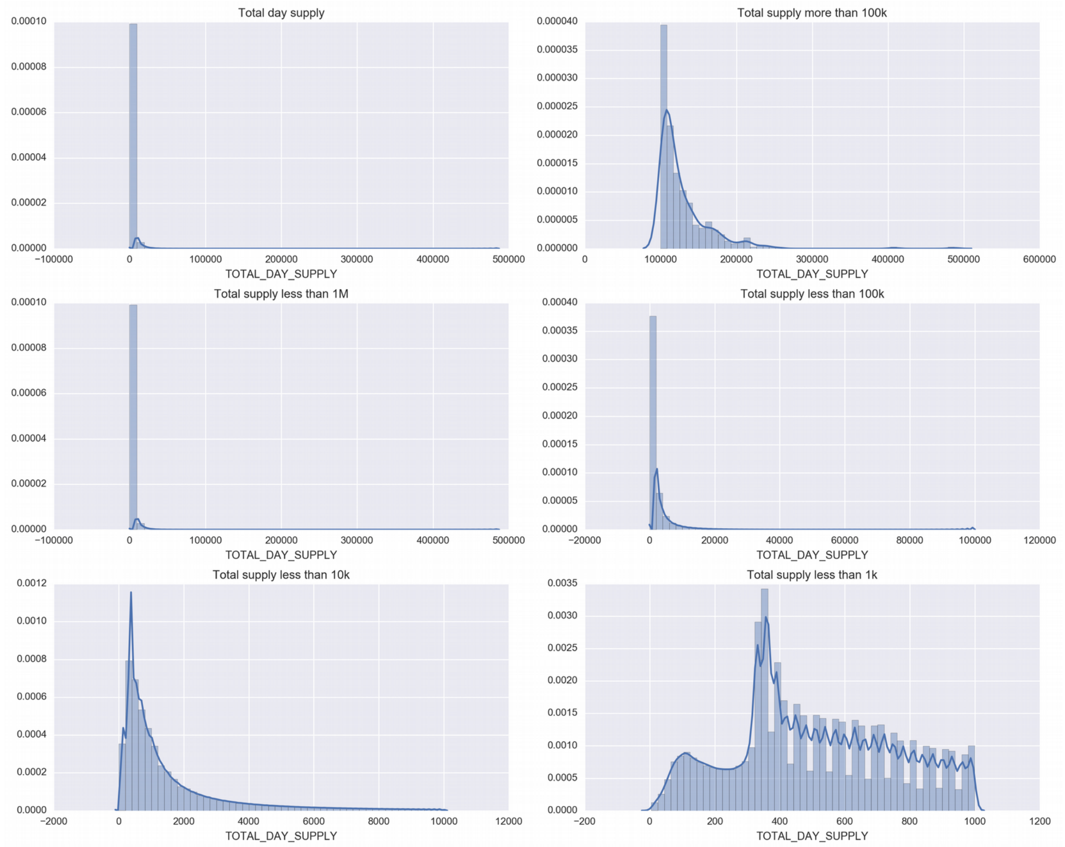


Exploration

- **Most physicians received less than $1,000 in 2014**
- **118 physicians received more than $1 Million**
- **Top earning physician**
  - $4.3 Million
  - Just from drug and device manufacturers!
  - Majority is from royalties and speaking fees
- **Reminder**
  - Payments != malice

Exploration

- **Claims data – Total day supply**
  - Total number of day's supply prescribed for each drug



Prediction Model

- **Random Forest**
- **Why?**
  - Linear models were not good
  - Random Forest showed a moderate level of predictability
  - Can interpret feature importance
  - Lots of categorical features
    - Physician Specialty
    - Drug Name

Prediction Model

- **Two models**
  - Both include the drug name and provider specialty
    - One-hot encoded
  - Both are predicting the total day supply
  - Only one had payments data
- **Outliers were included**
  - Hard to say at the outset if they should be excluded
- **Parameters cross-validated**
  - 140 trees, full growth, examine all possible splits

Results

Model	Train R-sq	Test R-sq	Train MSE	Test MSE	Train Mean AE	Test Mean AE	Train Median AE	Test Median AE
No payment information	0.359	0.358	1.09e7	1.08e7	1516.98	1519.26	650.01	652.79
Payment Information	0.907	0.357	1.57e6	1.09e7	566.78	1516.56	222.12	604.1

- **Model with payments much better training accuracy**
  - Artifact of having so much unique data
- **Testing median absolute error was 7.5% better with payments**
- **High mean errors – outliers are a problem**

Results

- **Feature importance from both**

No payments

	feature	importance
1349	LEVOTHYROXINE SODIUM	0.090877
2123	SIMVASTATIN	0.085321
226	AMLODIPINE BESYLATE	0.078578
1368	LISINOPRIL	0.078518
299	ATORVASTATIN CALCIUM	0.076280

With payments

	feature	importance
3612	Novo Nordisk Inc	0.045810
3355	Janssen Pharmaceuticals, Inc	0.035798
1349	LEVOTHYROXINE SODIUM	0.032142
2784	AstraZeneca Pharmaceuticals LP	0.030318
2123	SIMVASTATIN	0.029629

- **Drug manufacturer features were most important**
  - Indicates these manufacturers could be influencing prescription rates

Recommendations

- **Investigate business practices of manufacturers with highest feature importance**
- **Follow-up with patients of providers who have low prediction error rates**
- **Pending further investigation – Limit types and amount of payments made by drug manufacturers**

Considerations

- **Prediction was okay, large error mostly from outliers**
- **Payment forest was trained on 15% of data**
  - Took 7 hours on AWS EC2 r4.x16large (480GB, 64vCPU)
- **Could improve by**
  - Adding region information
  - Training on more data
  - Removing outliers
  - Associate payments only with drugs that are connected with the manufacturer

End

- **Special thanks to my mentor Andrew Vaughn!**