Optimizing Dillard's Expansion: Using Sales Data and Consulting-Backed Estimation to Select the Next Local Store Location(s)

•••

MLDS 400 Group 5
Xiyi Lin, Omar Shatrat, Fanqi Song, Tianyu Wu

Outline

- Overview of Dillard's
- Motivation & Business Question
- Our Data Science Pipeline
  - Data Cleansing and Exploratory Data Analysis (EDA)
  - Feature Selection and Engineering
  - Model Development and Selection
  - Recommendation - Decision Region
- ROI Spreadsheet and Analysis

Overview of Dillard's

- American department store chain
- Found in **1938**
- Operated over **300** Dillard's locations **as of 2005**
- **Online Store** at Dillars.com complements the physical retail spaces

Dillard's

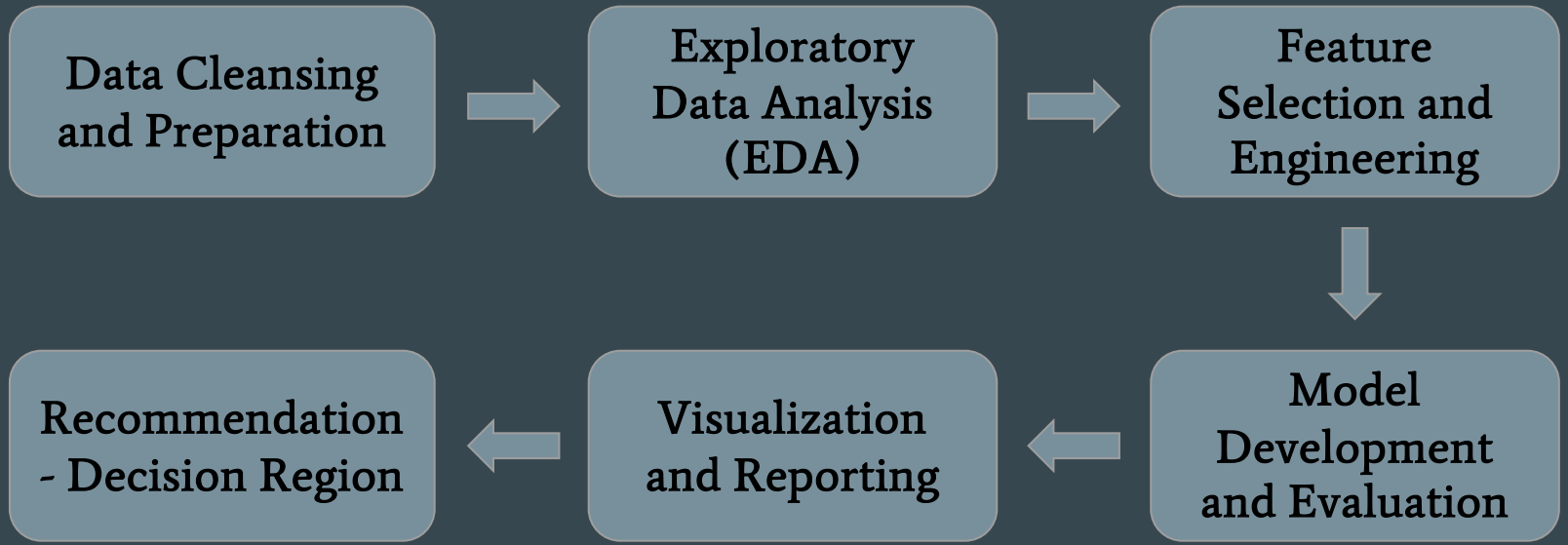
The Style of Your Life.

Motivation & Business Question

*Identify **promising location(s)** for Dillard's next local store by leveraging sales data and consulting-backed estimation*

- **Why?** Poor store location choices directly impact a store's long-term success and lead to operational inefficiencies affecting overall profitability
- **Our Goal**: Minimize risks in new store opening and ensure investments are directed for optimal returns

Our Data Science Pipeline

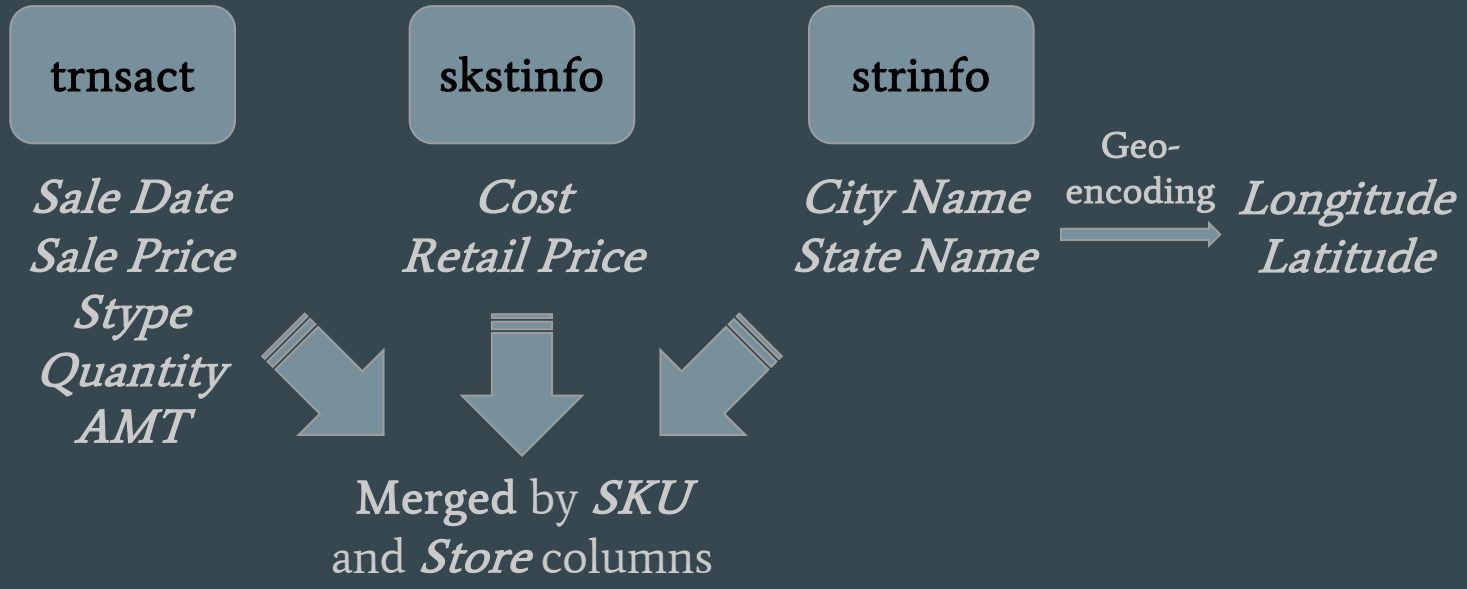


Data Science Pipeline - Data Cleansing

- **Missing values Imputation**
  - Typically in *cost*, *retail price*, *sale price*, and *AMT* columns
  - Fill by the mathematical relationship: *AMT* = *sale price* * *quantity*
  - Fill by average value: *Cost* <- groupby('SKU')['COST'].transform('mean')
- **Capture temporal patterns**
  - Fetch 'YYYY-MM' via *Sale Date* column
- **Geo-encoding**
  - using *city* and *state* columns to generate *latitude & longitude* columns

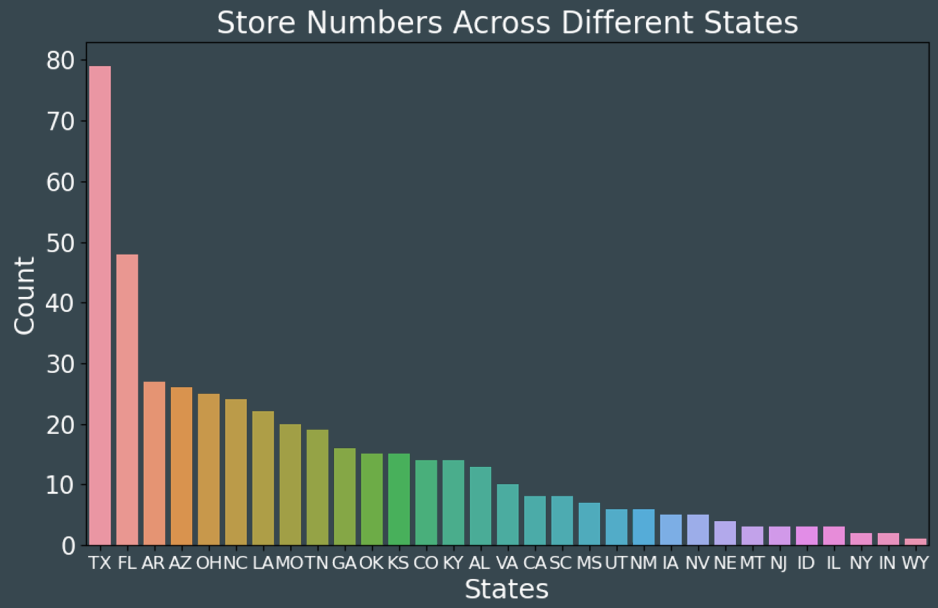
Data Science Pipeline - Data Preparation

- Tables of Interest
  - Ensure a comprehensive view of transaction, SKU, and store information



EDA: Store Location Distribution

- Stores across **30 states and 299 cities**

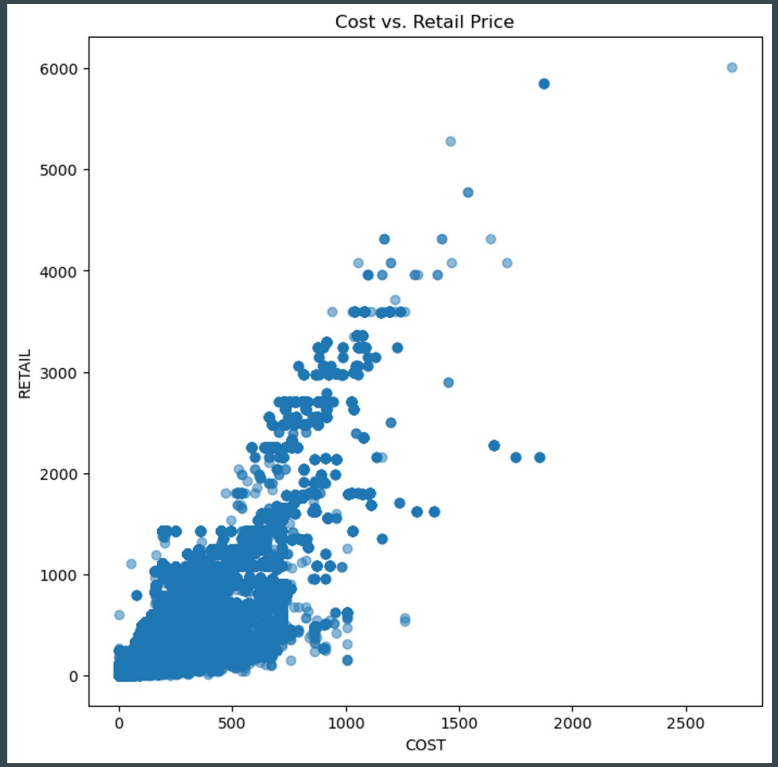


Number of Stores	
Top 5 by States	**Top 5 by Cities**
Texas (TX)	Little Rock **(AR)**
Florida (FL)	Gilbert **(AZ)**
Arkansas (AR)	Olathe **(KS)**
Arizona (AZ)	San Antonio **(TX)**
Ohio (OH)	Houston **(TX)**

Plausible Reason:
Heightened consumer demand and meticulous strategic planning

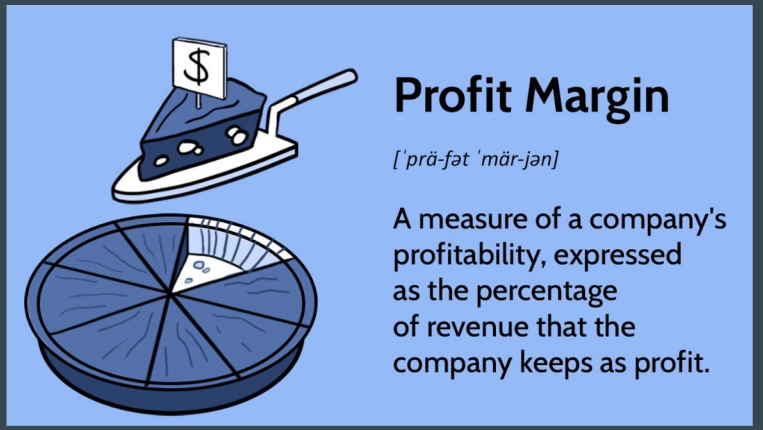
EDA: Cost and Retail Price

- A well-balanced distribution: moderate cost and retail price values
- Correlation (0.896): higher product costs correlate with elevated retail prices
- Insight: Having formulated effective pricing strategies



EDA: Profit Margins (PM)

- A metric that delineates the disparity between retail price and cost
- PM = (Retail Price - Cost) / Cost
- Insight: a significant store presence, efficient cost management, and well-executed pricing strategies



Top 5 Cities with the highest PM
Arkansas **(AR)**
Oklahoma **(OK)**
Ohio (OH)
Texas **(TX)**
Tennessee (TN)

Feature Selection and Engineering: KPIs

- **Average Cost of Goods Sold (COGS) per Item Sold**
  Indicating profitability by balancing sales with production costs. Lower values signify efficient sales generation with minimized production expenses.
- **MoM Growth**
  Calculating the percentage change in gross profit from the previous month. A positive MoM Gross Profit Growth indicates an increase in gross profit.
- **Return Percentage**
  Expressing the ***proportion of sales revenue returned*** to the business as product returns. A higher percentage implies a larger share of sales being returned.

Feature Engineering: KPIs

	STORE	AvgCOGS_peritem	MoMGrowth	ReturnPercentage	CITY
0	102	18.656348	0.102883	7.297510	TAMPA
1	103	17.761179	-1.857504	8.766210	ST LOUIS
2	107	16.975893	0.350877	8.382016	HURST
3	202	15.533738	0.190422	8.271629	TAMPA
4	203	16.687797	-0.614171	10.119624	CHESTERFIELD
...	...	...	...	...	...
326	9709	13.221052	-1.186029	6.835200	GREELEY
327	9804	14.891369	0.068109	7.260616	LAWTON
328	9806	17.679472	0.040775	4.250364	MABELVALE
329	9906	5.600000	NaN	0.000000	LITTLE ROCK
330	9909	13.783600	-0.668611	5.773841	CHEYENNE

Demo of KPI Values

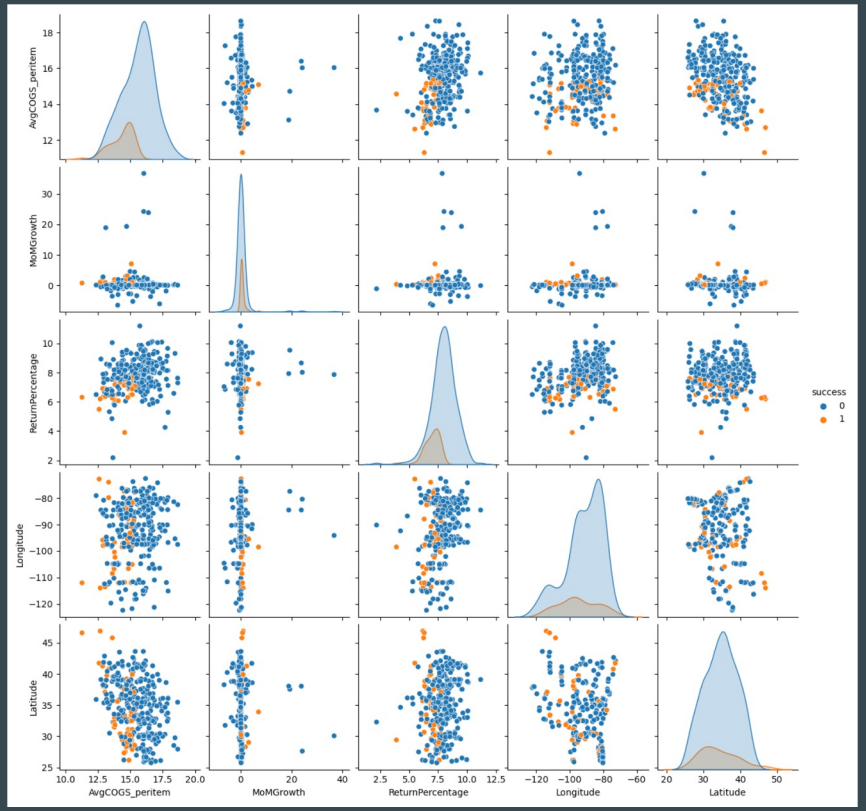
success
0
0
0
0
1
...
0
0
0
0
0

Feature Engineering Visualization

But ... how to define success?

Store is **successful** if and only if:

- *MoM Growth* is among the high-spread interval, i.e., above the median.
- *Average Cost of Goods Sold (COGS) per Item Sold* and *Return Percentage* is among the low-spread interval, i.e. below the median.



Pairplot of newly created features (KPIs)

Attention!

Irregular Correlation Heatmap

how?

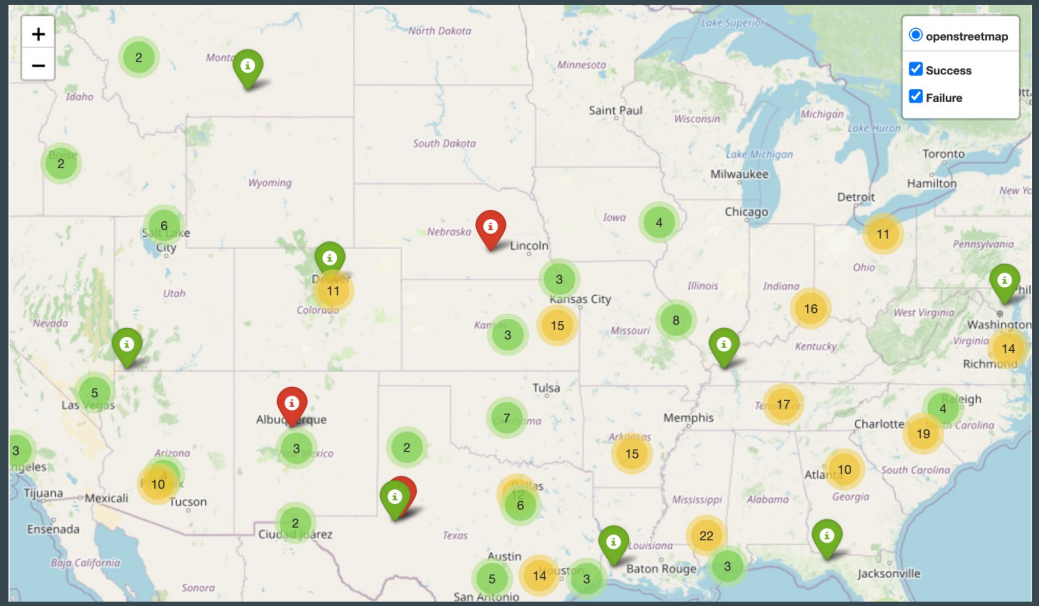
Our Decision:
Drop *Average Cost of Goods Sold (COGS) per Item Sold* feature when developing models

Successful Stores over the States

Now the success rate distribution is hence calculated as follows:

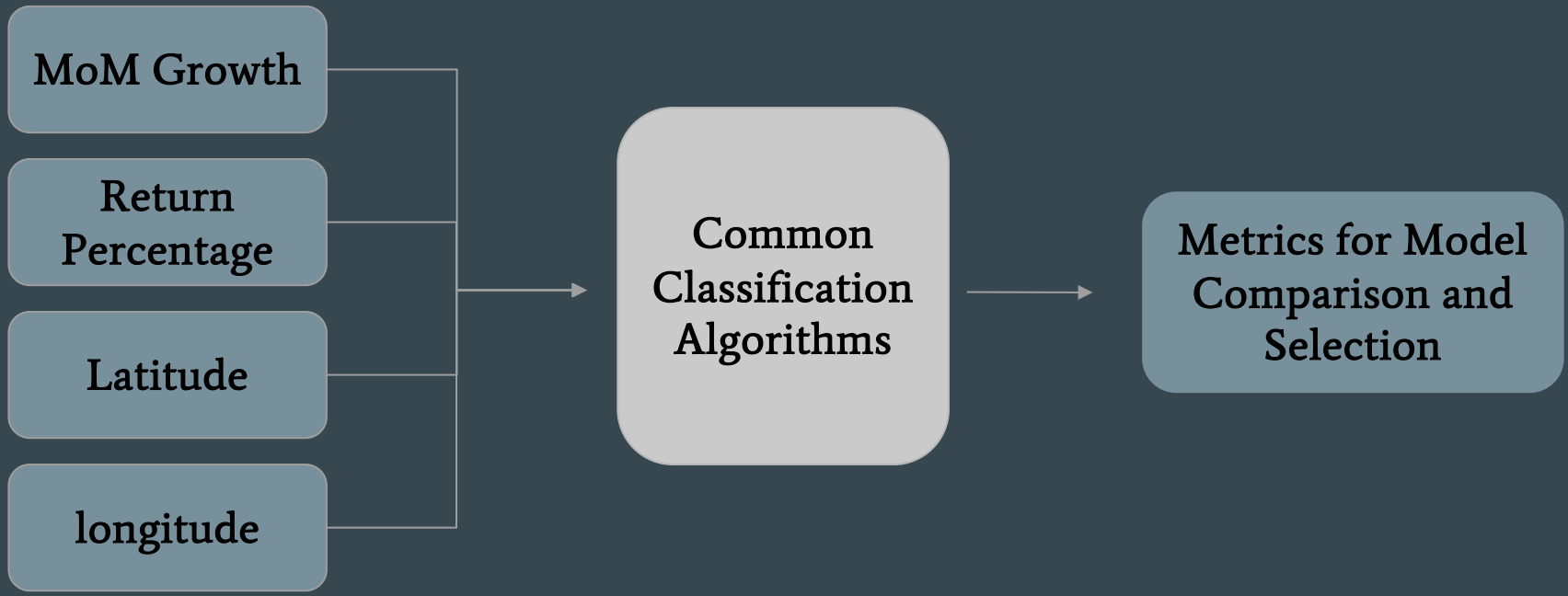
Successful	Not Successful
15.3%	84.7%

=> The ratio between successful and unsuccessful stores appears to be **reasonable**



Demo of Store Success Distribution On Map

Model Development



Model Comparison & Selection

Models	Metrics			
	Accuracy	Macro Precision	Macro Recall	Macro F-1 Score
Logistic Regression	0.76	0.64	0.77	0.65
Decision Tree	0.89	0.78	0.75	0.76
Random Forest	0.91	0.80	0.85	0.82
SVM	0.65	0.56	0.61	0.54
K-NN	0.70	0.62	0.73	0.60

=> We will then use **Random Forest** model to select the *promising location(s)* for Dillard's next local store.

Recommendation - Decision Region

Based on consulting-backed estimations, our model shows that **COLUMBIA** and **PHOENIX** emerge as the most promising locations among the five of interest below

	MoMGrowth	ReturnPercentage	Latitude	Longitude	CITY	Success
1	20.4012	0.002	35.395	-95.814	OKLAHOMA	0
2	0.8123	4.28038	39.204	-76.690	COLUMBIA	1
3	-0.12595	10.2903	32.380	-86.312	MONTGOMERY	0
4	0.067734	6.32	33.415	-111.835	MESA	0
5	0.4729	5.2039	33.451	-112.016	PHOENIX	1

ROI Spreadsheet

US Clothing Market Size	$	351,400,000,000.00	
CAGR		1.93%	
Proj. Annual Mkt.	$	358,182,020,000.00	
Dillards 2022 Revenue	$	6,900,000,000.00	
Dillards 2022 OpEx	$	1,674,000,000.00	
# of Dillards Stores		277	
Revenue / location	$	24,909,747.29	
OpEx / location	$	6,043,321.30	
Margin / location	$	18,866,425.99	
Planned Expansion		5.3461	
Fixed opening costs	$	100,000.00	
Rent / sq ft.	$	61.40	
Typical sq. ft. of store		250,000.00	
Yearly Lease	$	15,350,000.00	
Annual Profit / store	$	3,416,425.99	
% Stores Successful		15.30%	from model
Proj. Successful Stores		0.8179533	
ROI	$	2,794,476.92	

ROI Analysis

- The **market opportunity** for Dillard's is **significant**:
  - **Total Market for Clothing: $351.4bn**
  - **Annual Avg. Growth: 1.93%**
- Being able to assess which stores will be strong performers can help Dillard's cut down on fixed costs and lease expenses.
- If Dillard's grows at a similar pace with the market, ROIs in the neighborhood of **~$2.8bn** are attainable.

Thanks for Listening!