Customer segmentation for credit card company

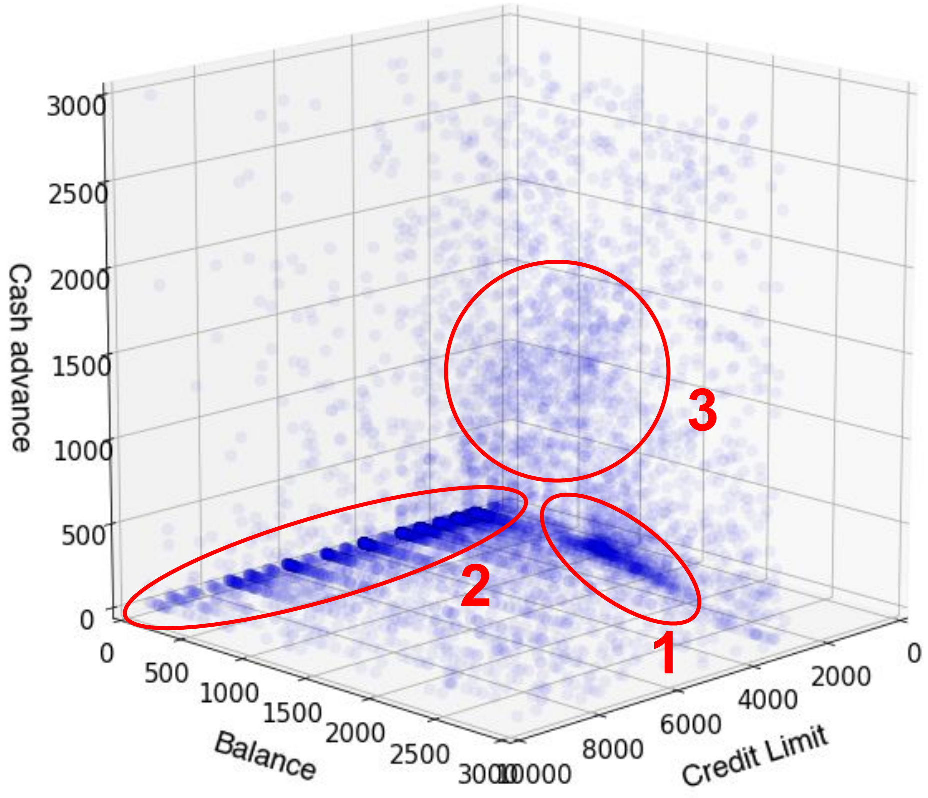
Michael A. Boles
Insight Data Science
Data Challenge 2
February 28, 2020

Credit card user segmentation strategy

Executive summary

Three key user segments:

1. Heavy credit users (Balance ≈ Credit limit)
   - Target these customers by offering competitive (high) credit limit
   - But pay attention to credit risk
2. Light credit users (Balance << Credit limit)
   - Keep these customers engaged (spending) by offering cash back
   - But avoid excessive generosity
3. Cash users (Cash advance > 0)
   - Target these customers with low cash advance fees
   - But pay attention to credit risk



Presentation outline

Supplementary slides

1. Exploratory data analysis

- Number of entries: 8247
- Number of features: 17
- Data types: integers/floats
- Correlation: mostly weak

Slides 4 - 6

2. Dimensionality reduction and modeling

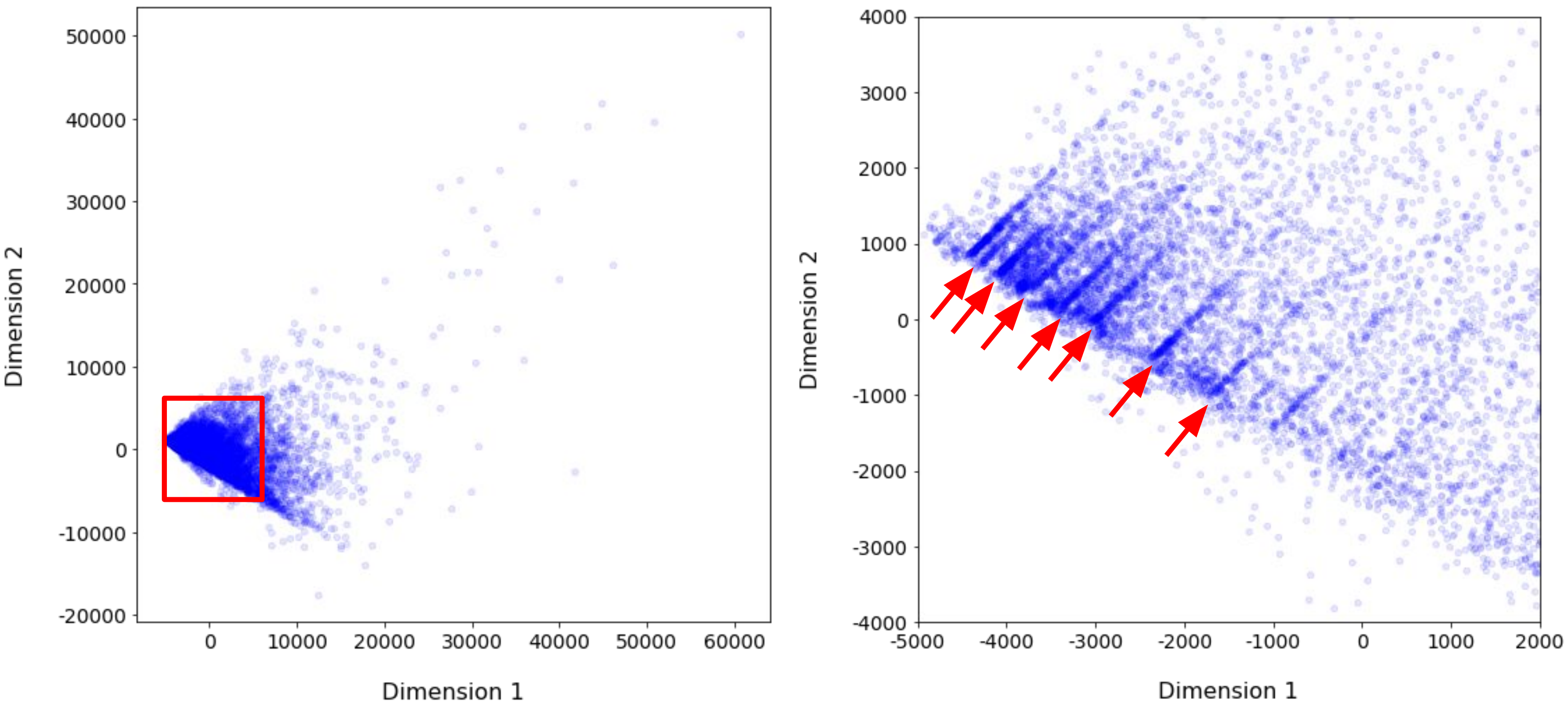
- Reduce 17 -> 2 dimensions (PCA)
- Examine feature contributions to PCA axes
- Clustering (k-means, Gaussian mixture)

Slides 7, 8

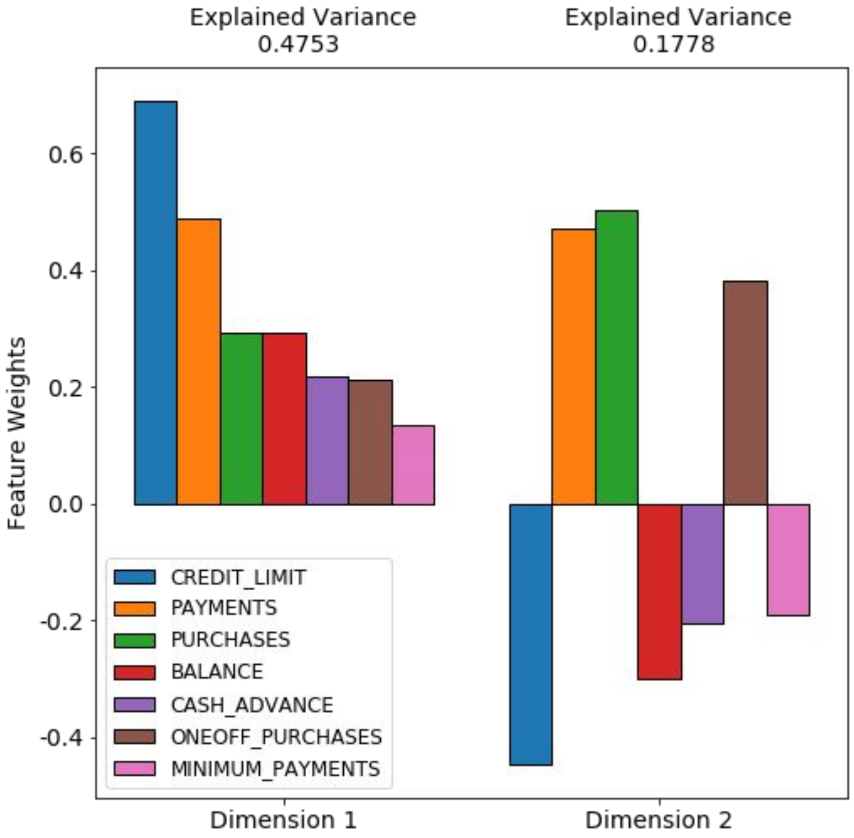
3. Customer segmentation

- Compare cluster orientation vs. eigenvectors of most important features
- Visualize scatter data along 3 most important axes
- Determine customer segments

PCA hints at grouping of customer behaviors



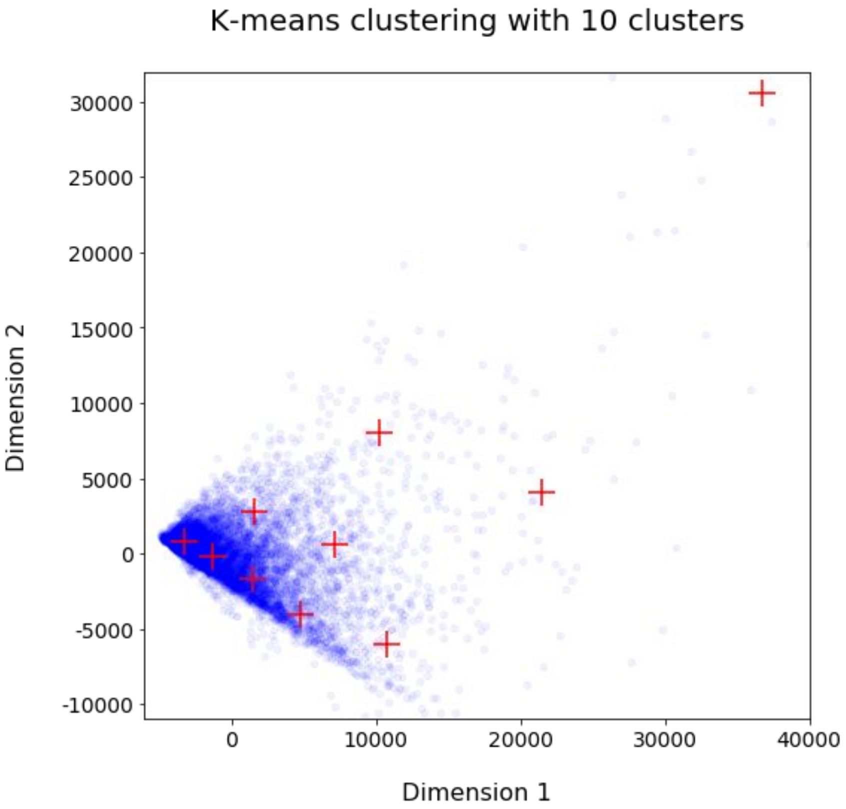
PCA weights show features with most variance

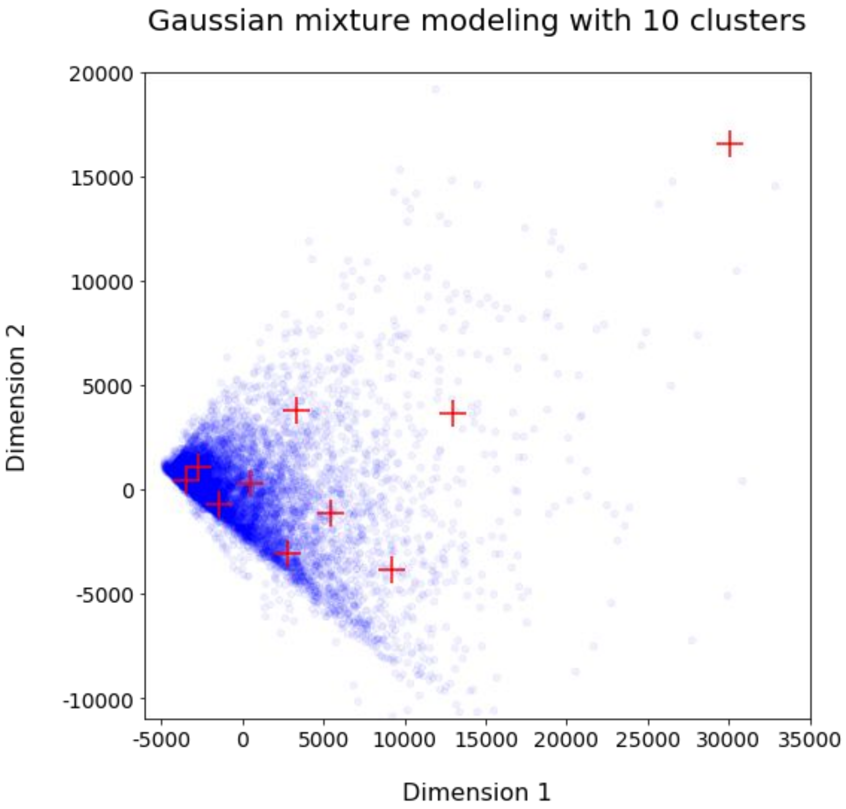


	Dimension 1	Dimension 2
CREDIT_LIMIT	0.6901	-0.4474
PAYMENTS	0.4874	0.4703
* **PURCHASES**	0.2927	0.5008
BALANCE	0.2915	-0.3007
CASH_ADVANCE	0.2163	-0.2044
ONEOFF_PURCHASES	0.2108	0.3804
MINIMUM_PAYMENTS	0.1325	-0.1910
INSTALLMENTS_PURCHASES	0.0818	0.1205
PURCHASES_TRX	0.0023	0.0035
CASH_ADVANCE_TRX	0.0004	-0.0005

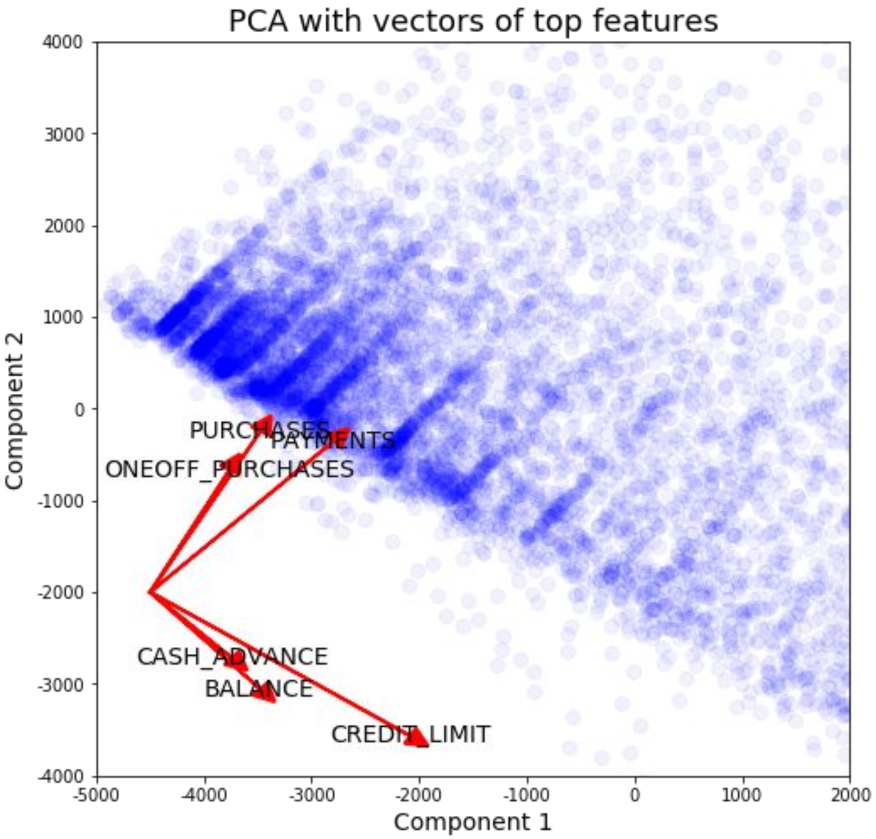
Features most capable of distinguishing differences in user behavior

Clustering: no clear correlation between cluster center assignment and scatter data





Customer segments separated by 3 features



Scatter data

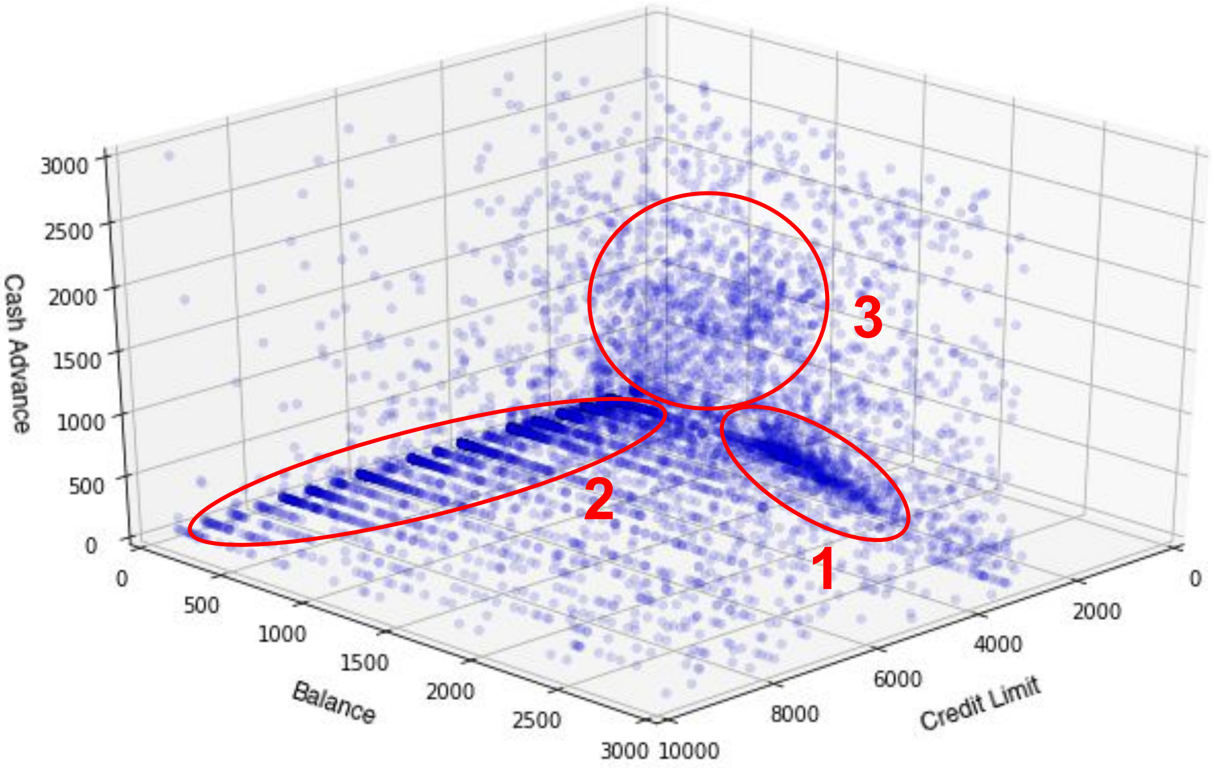
- Points are clustered along lines from bottom left to top right (positive slope)
- These striations are separated from one another along the orthogonal direction

PCA eigenvectors

- Credit limit, Balance, and Cash advance features are perpendicular to clusters
- Payments, Purchases, and One-off purchases features are parallel to clusters

Features to best distinguish customer segments:
Credit limit, Balance, and Cash advance

There are 3 key types of credit card users



Customer segments

1. *Heavy credit users*
   (Balance ≈ Credit limit)

2. *Light credit users*
   (Balance << Credit limit)

3. *Cash users*
   (Cash advance > 0)

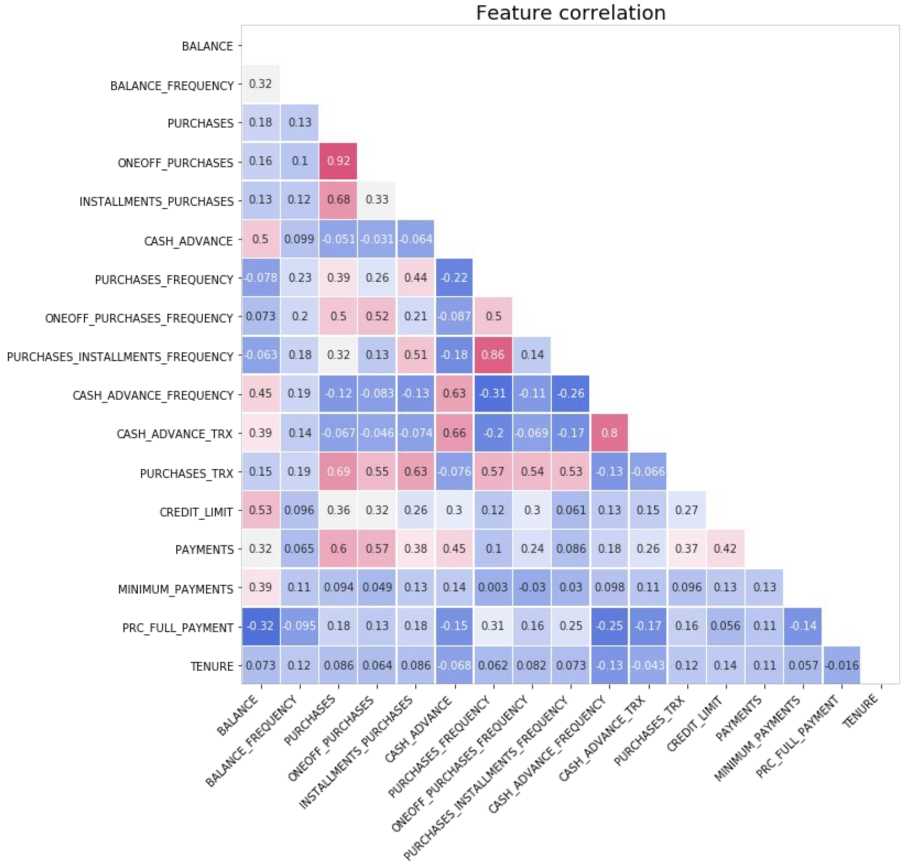
End

Supplementary slide: exploratory analysis

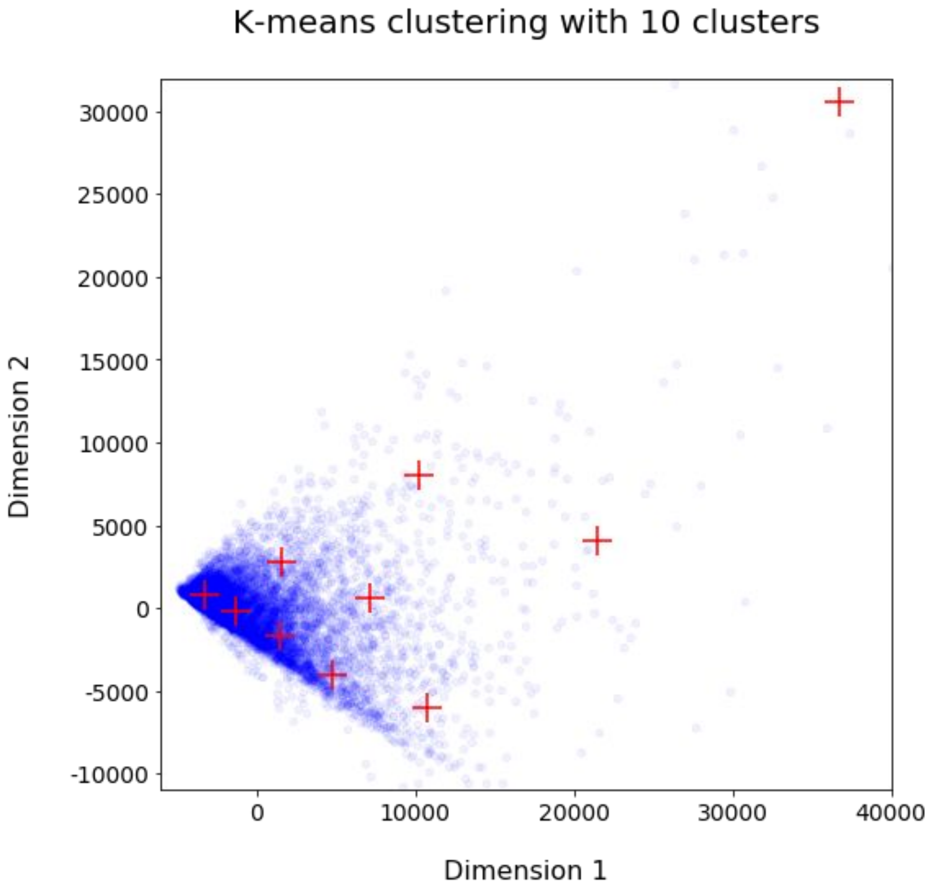
	count	mean	std	min	25%	50%	75%	max
BALANCE	8247.0	1381.584865	1739.703180	0.000000	115.774483	810.089776	1864.398128	9749.239122
BALANCE_FREQUENCY	8247.0	0.875091	0.239780	0.000000	0.888889	1.000000	1.000000	1.000000
PURCHASES	8247.0	801.395230	1160.706492	0.000000	43.225000	354.000000	1045.355000	9030.080000
ONEOFF_PURCHASES	8247.0	462.320187	894.775671	0.000000	0.000000	34.320000	537.105000	7025.020000
INSTALLMENTS_PURCHASES	8247.0	339.313636	559.771284	0.000000	0.000000	85.840000	449.260000	4019.990000
CASH_ADVANCE	8247.0	786.845310	1452.936266	0.000000	0.000000	0.000000	987.224393	9337.271222
PURCHASES_FREQUENCY	8247.0	0.485025	0.398764	0.000000	0.083333	0.500000	0.916667	1.000000
ONEOFF_PURCHASES_FREQUENCY	8247.0	0.194807	0.291026	0.000000	0.000000	0.083333	0.250000	1.000000
PURCHASES_INSTALLMENTS_FREQUENCY	8247.0	0.358833	0.394468	0.000000	0.000000	0.166667	0.750000	1.000000
CASH_ADVANCE_FREQUENCY	8247.0	0.122829	0.180376	0.000000	0.000000	0.000000	0.166667	0.916667
CASH_ADVANCE_TRX	8247.0	2.691767	4.731430	0.000000	0.000000	0.000000	4.000000	30.000000
PURCHASES_TRX	8247.0	12.672972	17.192131	0.000000	1.000000	7.000000	16.000000	114.000000
CREDIT_LIMIT	8247.0	4215.685477	3242.614935	50.000000	1500.000000	3000.000000	6000.000000	18500.000000
PAYMENTS	8247.0	1391.383457	1641.293154	0.000000	381.419802	812.070417	1730.785825	12902.188130
MINIMUM_PAYMENTS	8247.0	650.109541	914.630857	0.019163	169.791896	316.230814	845.654374	10057.561920
PRC_FULL_PAYMENT	8247.0	0.150982	0.289034	0.000000	0.000000	0.000000	0.142857	1.000000
TENURE	8247.0	11.639505	1.063429	7.000000	12.000000	12.000000	12.000000	12.000000

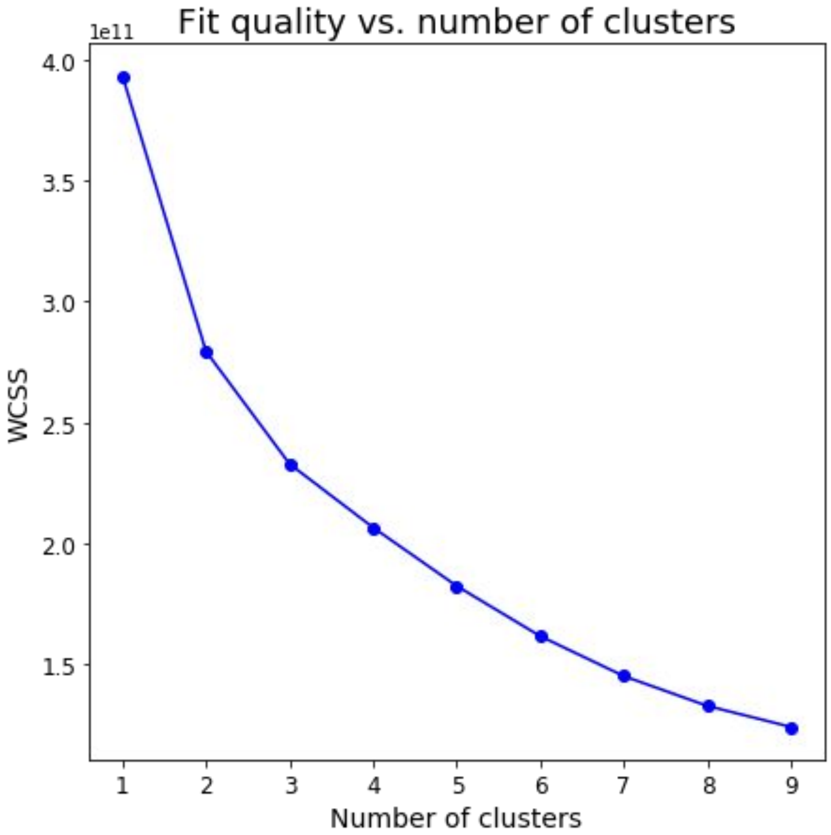
Supplementary slide: exploratory analysis

BALANCE
BALANCE_FREQUENCY
PURCHASES
ONEOFF_PURCHASES
INSTALLMENTS_PURCHASES
CASH_ADVANCE
PURCHASES_FREQUENCY
ONEOFF_PURCHASES_FREQUENCY
PURCHASES_INSTALLMENTS_FREQUENCY
CASH_ADVANCE_FREQUENCY
CASH_ADVANCE_TRX
PURCHASES_TRX
CREDIT_LIMIT
PAYMENTS
MINIMUM_PAYMENTS
PRC_FULL_PAYMENT
TENURE

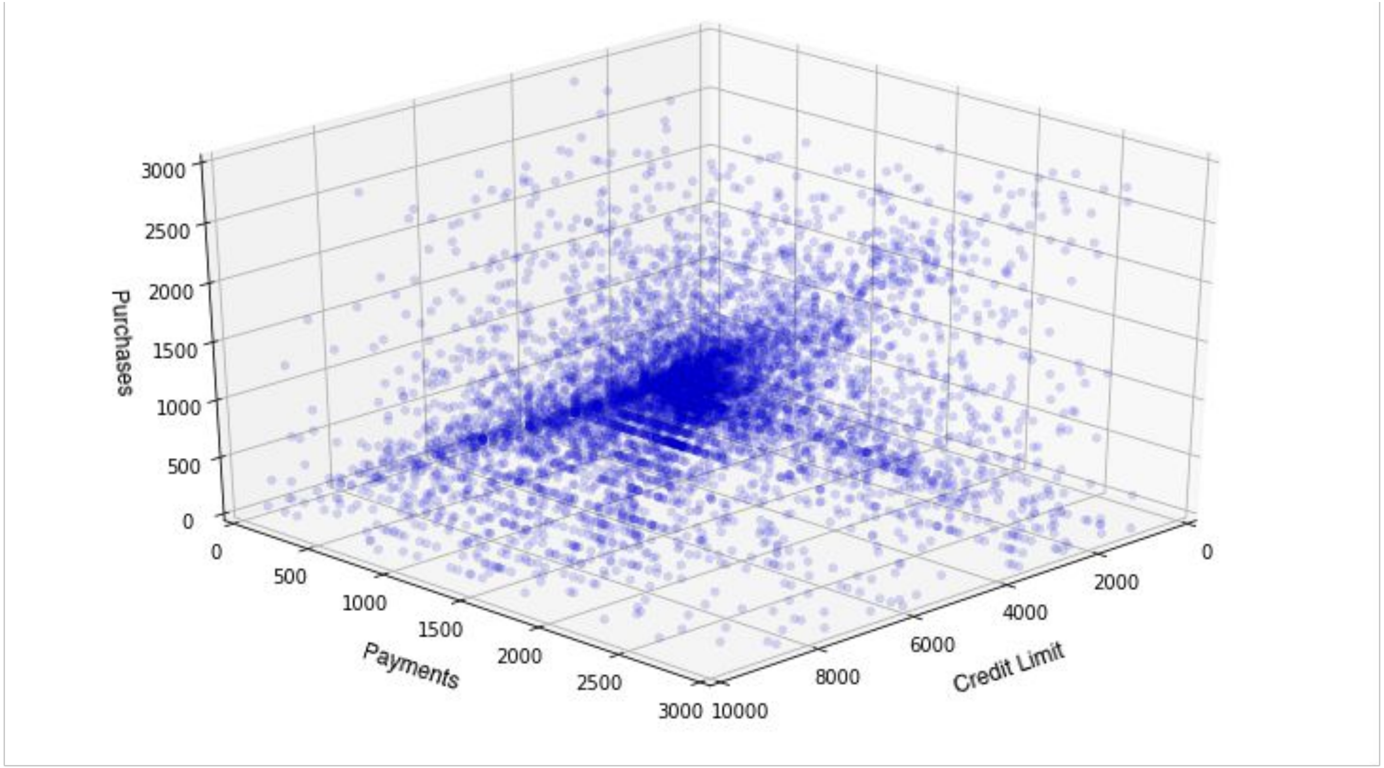


Supplementary slide: k-means clustering





Credit limit, payments, and purchases: less clear separation of scatter data



Credit card user segmentation

Prompt

Data science at credit card company
- Goal: reduce costs of offering sign-up incentives to new customers
- Currently giving everyone same incentives
- Wish to carefully target benefits to customer segments

Examine cardholder data
- Understand customer segments
- Which benefits might attract each segment?

Executive summary

Three key user segments
- Heavy credit users (Balance ≈ Credit limit)
  - Target these customers by offering competitive (high) credit limit
  - But pay attention to credit risk
- Light credit users (Balance << Credit limit)
  - Keep these customers engaged (spending) by offering cash back
  - But avoid excessive generosity
- Cash users (Cash advance > 0)
  - Target these customers with low cash advance fees
  - But pay attention to credit risk