Master the Mainframe World Championship 2016

Apache Spark on IBM z/OS

Stephen Solis-Reyes

The University of Western Ontario
London, Ontario, Canada

September 17, 2016

Outline

1. About Spark (what I learned, what I liked)
2. About the data
3. My analysis/insights

About Spark

Spark is a platform for large-scale data analytics.

About Spark

Spark is a platform for large-scale data analytics.

- Multiple languages: Scala, Python, Java
- Large variety of input data sources
- Several libraries:
  - MLlib (machine learning)
  - GraphX (graph processing)
  - Spark Streaming (near-real-time streaming analytics)

What I liked about Spark

- Extremely scalable, automatically (10000s of nodes, PBs of data)

What I liked about Spark

- Extremely scalable, automatically
  (10000s of nodes, PBs of data)
- Very fault-tolerant
  - on failure, work is moved to another worker automatically – no data loss

What I liked about Spark

- Extremely scalable, automatically (10000s of nodes, PBs of data)
- Very fault-tolerant
  - on failure, work is moved to another worker automatically – no data loss
- Cross-platform, runs anywhere Java runs *without code changes*

What I liked about Spark

- Extremely scalable, automatically (10000s of nodes, PBs of data)
- Very fault-tolerant
  - on failure, work is moved to another worker automatically – no data loss
- Cross-platform, runs anywhere Java runs *without code changes*
- Interactive shell
  - can easily explore data ‘live’, no need to compile and wait

Spark on z/OS

- Run analytics right where the data is
  - less network use, lower latency, security, etc.

Spark on z/OS

- Run analytics right where the data is
    - less network use, lower latency, security, etc.
- Can use many z/OS-native data sources
    - DB2, VSAM, IMS, PDSE, etc.

Spark on z/OS

- Run analytics right where the data is
  - less network use, lower latency, security, etc.
- Can use many z/OS-native data sources
  - DB2, VSAM, IMS, PDSE, etc.

Contest system

z13 running z/OS, data stored in DB2

Dataset

~6000 clients

~1.5M transactions
(year 2013, ~$75.2M total)

Clients

Gender (A/B)

Age

Annual income

Education level

Service discontinued

Transactions

Client

Date and Time

Amount

Merchant name/category

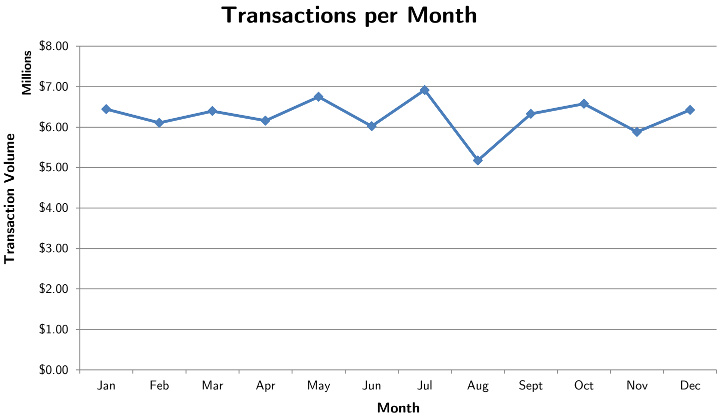
Card type/brand

About the Analysis

Predictive analytics drives smart business decisions

- What actionable business insights can we draw from the data, using Spark?

Transaction Behaviour Over Time

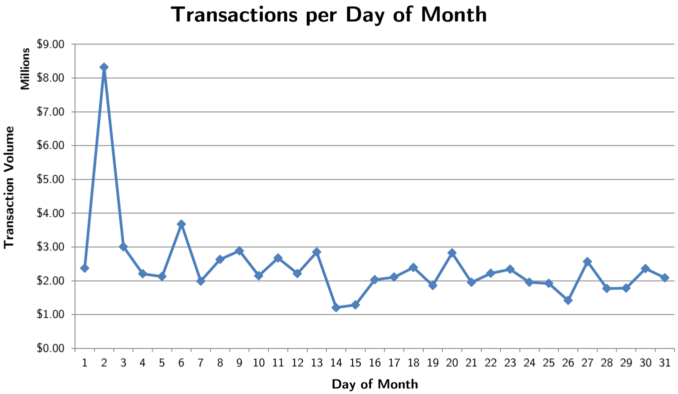


Business Insights

Transaction volume is fairly steady per month.

- Probably not much benefit in monthly card promotions

Transaction Behaviour Over Time

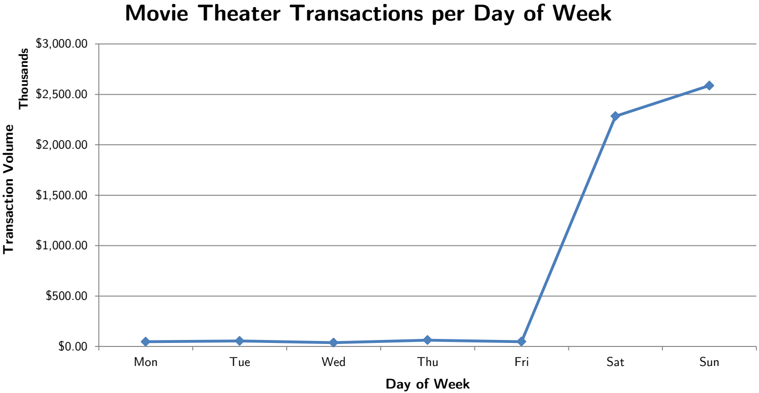


Business Insights

Huge spike on the 2nd of every month.

- Businesses should be prepared for the volume and use promotions to take advantage

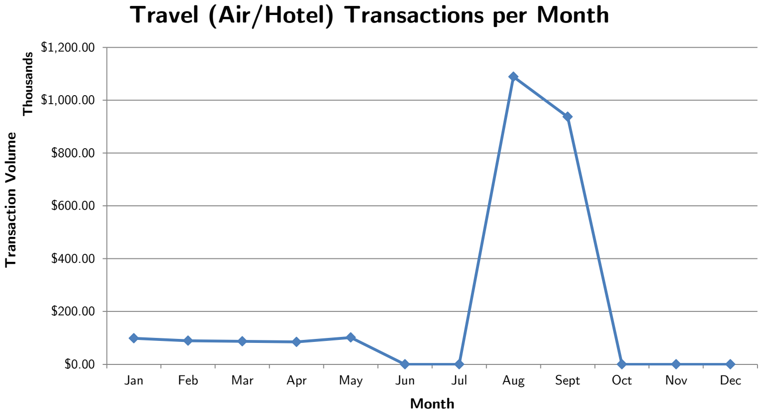
Merchant Category Insights



Business Insights

- Friday nights are surprisingly unpopular
- Weekday discounts (not just Tuesdays) could be effective

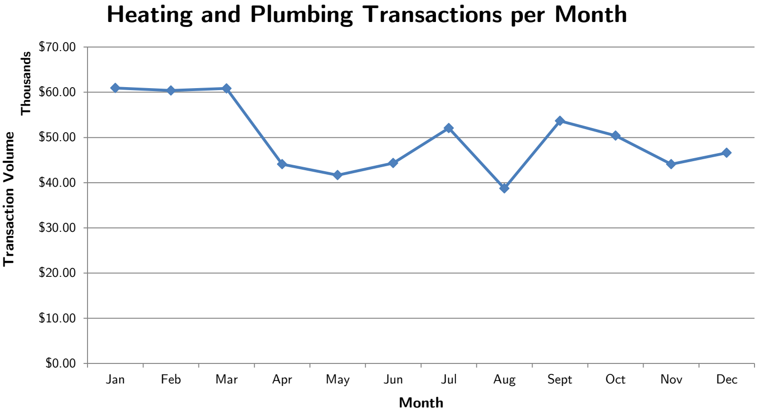
Merchant Category Insights



Business Insights

- Travel-related businesses should anticipate the spike in Aug/Sept
- Could offer discounts in Jul/July, Oct/Nov/Dec to encourage spending

Merchant Category Insights

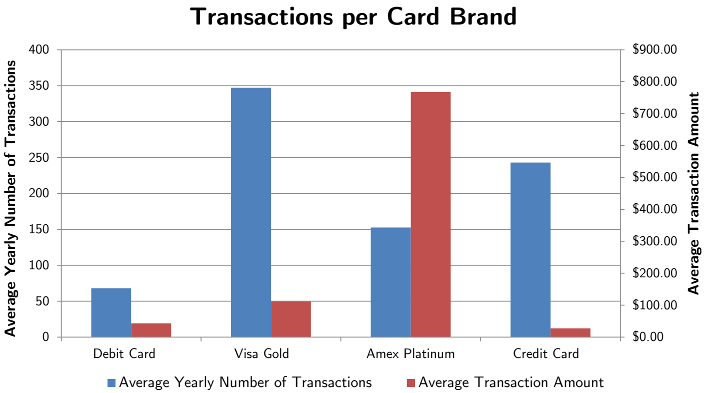


Business Insights

Definite monthly trends, perhaps caused by weather/climate.

- Could consider discounts in off periods, hire seasonal employees during high demand

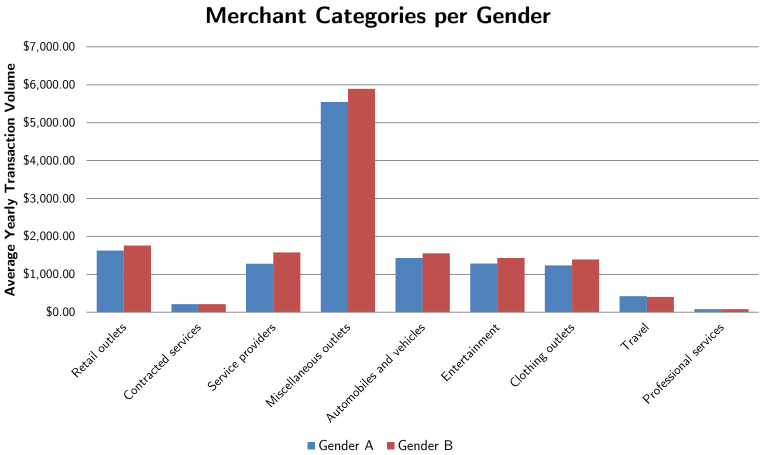
Card Brands



Business Insights

- Banks should push credit cards, since people make more transactions with them
- Platinum card customers make much larger transactions, could push more expensive promos

Gender

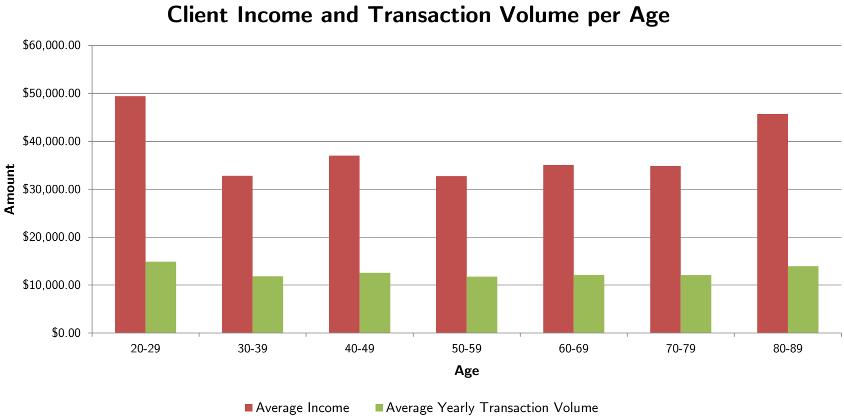


Business Insights

Different genders do not seem to have different spending patterns.

- Gender-specific marketing may not be so effective

Age



Business Insights

Young and old clients have more income, but do not spend more money.

- Age-specific marketing may not be so effective

> "Information is the oil of the 21st century, and analytics is the combustion engine."
>
> — Peter Sondergaard, Gartner Research

Thanks! Questions?

Me: Stephen Solis-Reyes