SJSU - CS185C - Big Data
Instructor:
Peter Zadrozny

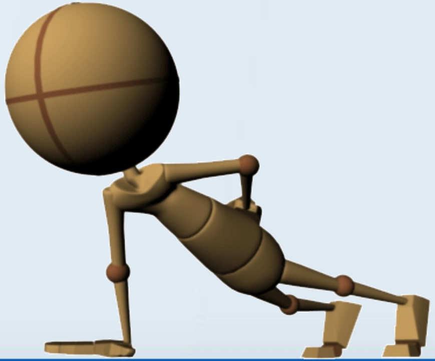
Hadoop- Airline On Time

Team # 8

Nam Phan
Mallika Perepa
Liu PingChuan

AGENDA

1. Briefly about the Cluster ...
2. What is the project about?
3. Storyline



1.Cluster and Project Details…..

About the Cluster -

- Created virtual servers using GoGrid account
- □Used Cloudera Distribution to install hadoop
- Created a cluster of 5 nodes:
  - 1 master node (capacity-8GB)
  - 4 slave nodes (capacity-4GB)

About the Project -

- Aim -
  - Download the data set
  - Create queries to interpret the data set
  - Visualize the data obtained by running the queries
- Data Set -
  - The data set comprises of scheduled vs actual and departure times for domestic flights by US certified carriers...

2. Project Data Set

- ➢ Downloaded the Airline On-Time data which is available at -
  - ○ http://www.transtats.bts.gov/DL_SelectFields.asp?Table_ID=236&DB_Short_Name=On-Time
  - ○ http://www.faa.gov/licenses_certificates/aircraft_certification/aircraft_registry/releasable_aircraft_download

- ➢ Size of the data set : 843 MB
  - ○ 1 Year of Data(August 2012 - July 2013)
  - ○ 6268846 Records

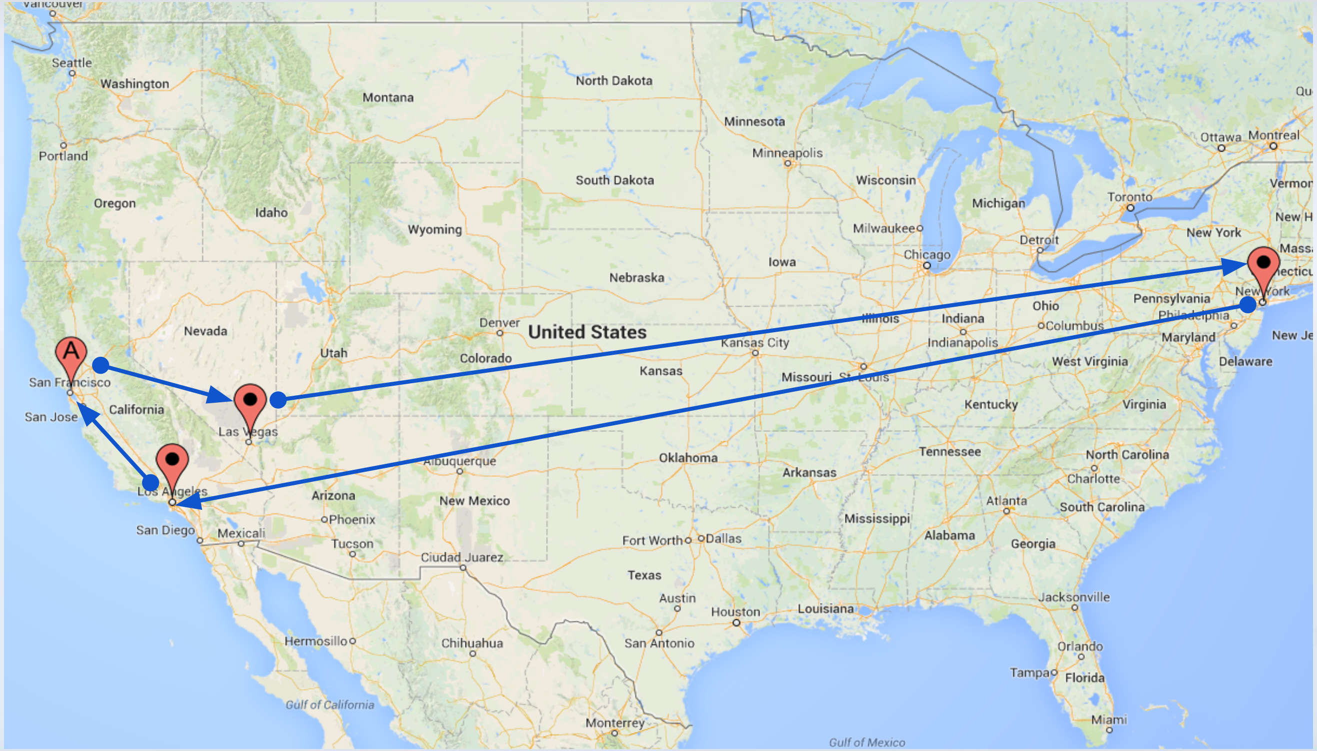
- ➢ Loaded the data set on the Hadoop cluster
  - ○ Created tables using Hive…
  - ○ Created queries to answer the metrics…
  - ○ Verified the output of the queries on small set of data...

3.Story

Our man and his story...

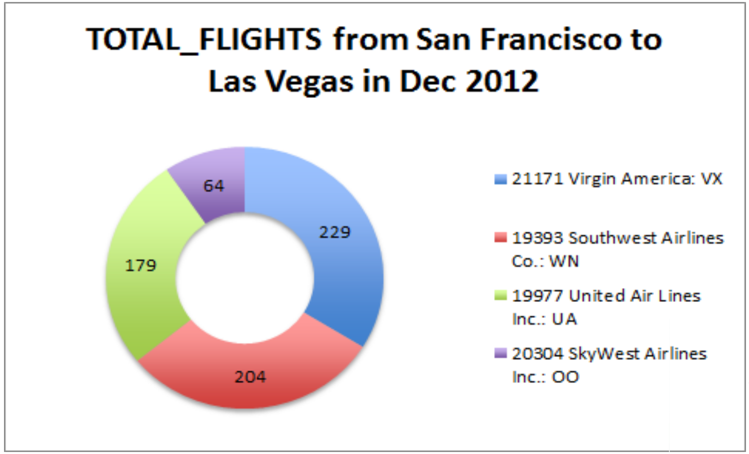
- Name: Joe
- Status: Single
- Plan: To take a vacation in December around the U.S.

Joe’s Trip

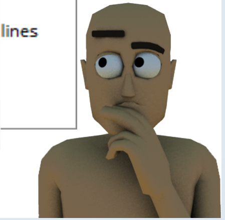

Vancouver
Seattle
Washington
Portland
Oregon
Idaho
Montana
North Dakota
Minnesota
Minneapolis
South Dakota
Wyoming
Wisconsin
Michigan
Milwaukee
Chicago
Detroit
Toronto
Ottawa
Montreal
Qu
Vermon
New H
New York
Mass
necticu
Iowa
Nebraska
Nevada
Utah
Denver
Colorado
United States
Kansas City
Kansas
Illinois
Indiana
Indianapolis
Ohio
Columbus
Pennsylvania
Philadelphia
New York
Maryland
New Je
Delaware
West Virginia
Missouri
St. Louis
Kentucky
Virginia
A
San Francisco
San Jose
California
Las Vegas
Los Angeles
San Diego
Mexicali
Arizona
Phoenix
Tucson
Albuquerque
New Mexico
Oklahoma
Arkansas
Tennessee
North Carolina
Charlotte
South Carolina
Atlanta
Mississippi
Alabama
Georgia
Fort Worth
Dallas
Texas
Ciudad Juarez
Austin
Houston
San Antonio
Louisiana
Jacksonville
Orlando
Tampa
Florida
Miami
Hermosillo
Chihuahua
Gulf of California
Monterrey
Gulf of Mexico

San Francisco to Las Vegas

- McCarran International Airport (Las Vegas)
- 4 airlines
- Joe randomly chooses **Southwest Airlines**



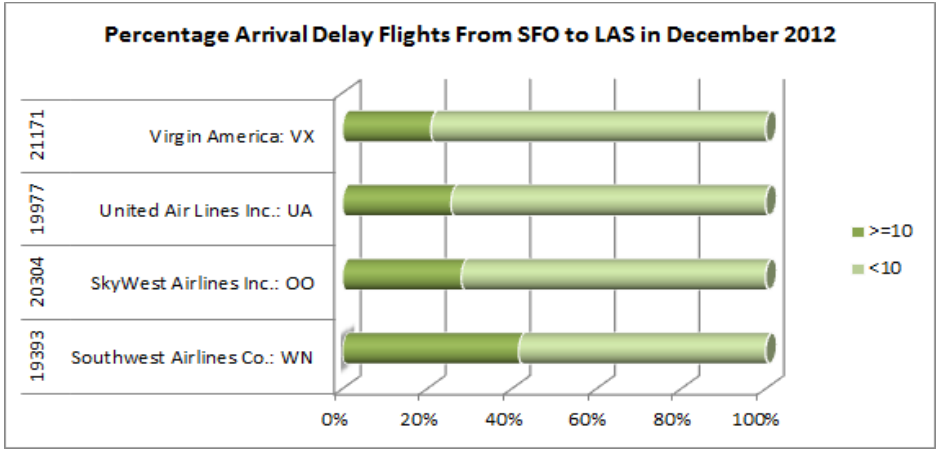
TOTAL_FLIGHTS from San Francisco to Las Vegas in Dec 2012



San Francisco to Las Vegas Contd….

- Joe’s flight gets **delayed** badly
- Why?

- Joe doesn’t know that 40% flights of Southwest Airlines are delayed

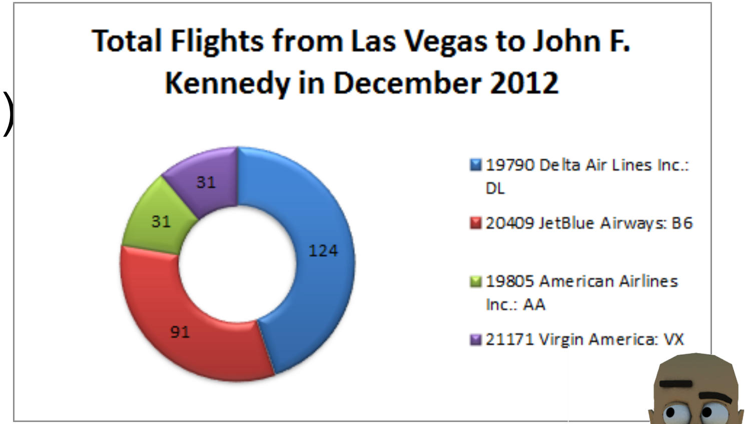


Percentage Arrival Delay Flights From SFO to LAS in December 2012

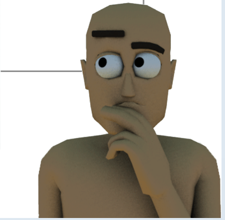
21171
Virgin America: VX
19977
United Air Lines Inc.: UA
20304
SkyWest Airlines Inc.: OO
19393
Southwest Airlines Co.: WN
0%
20%
40%
60%
80%
100%
>=10
<10

Las Vegas to New York

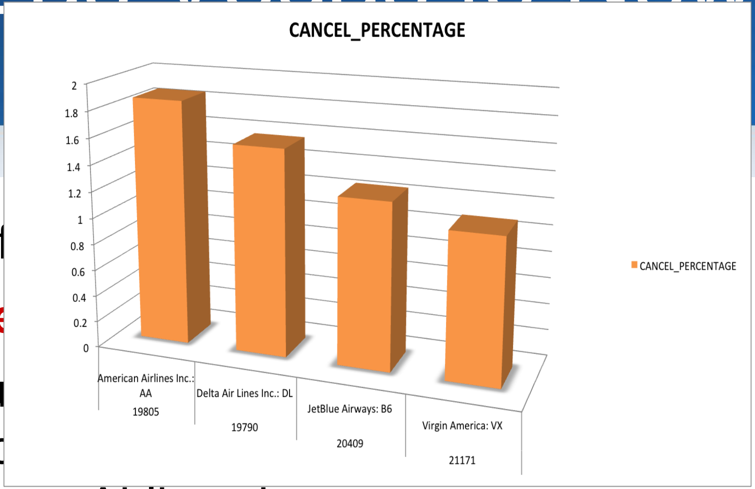
- ➢ John F. Kennedy International (New York)
- ➢ 4 airlines
- ➢ Joe randomly chooses American Airlines



Total Flights from Las Vegas to John F. Kennedy in December 2012



Las Vegas to New York Contd….



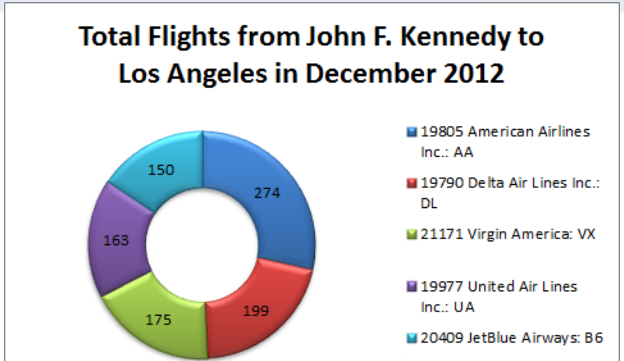
- Joe's f

  cance

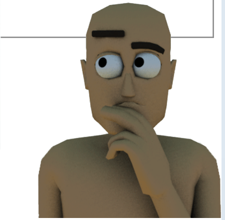
- How u
- Joe do
  American Airlines has the highest cancellation rate in the 4 airlines

New York to Los Angeles

- ➢ Los Angeles International Airport (Los Angeles)
- ➢ 5 airlines
- ➢ Joe doesn't want to randomly choose again

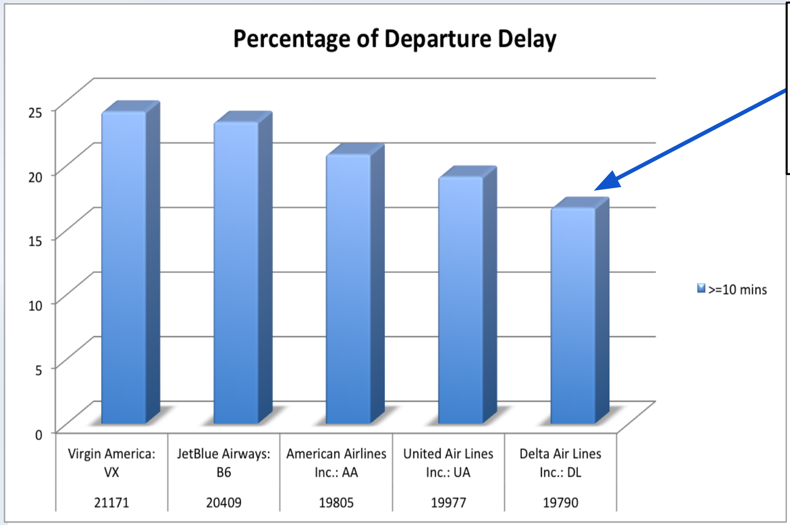


Total Flights from John F. Kennedy to Los Angeles in December 2012

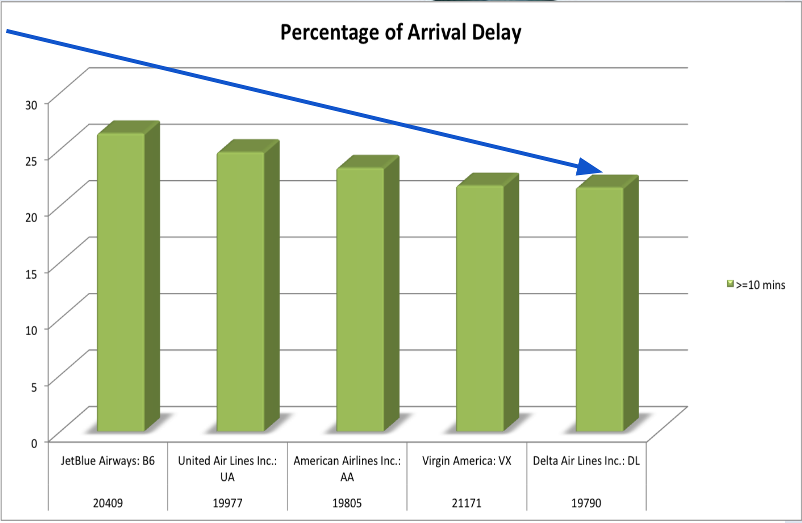


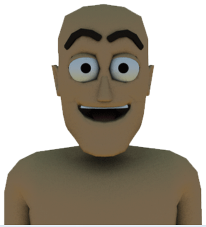
- ➢ Joe calls his friend for advising about airlines with lowest delay rates

New York to Los Angeles Contd..



Delta Air Lines is the best choice for Joe

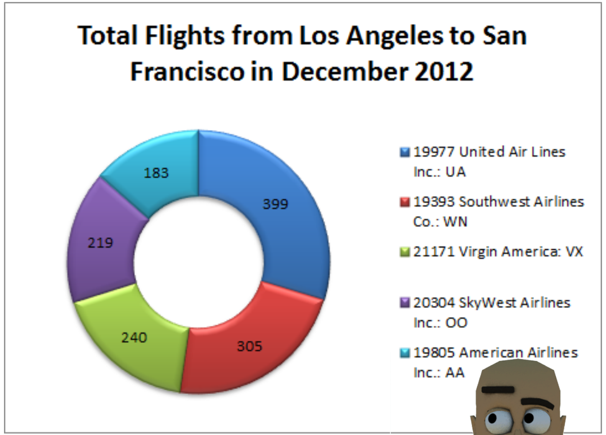


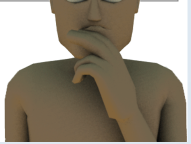


Los Angeles to San Francisco

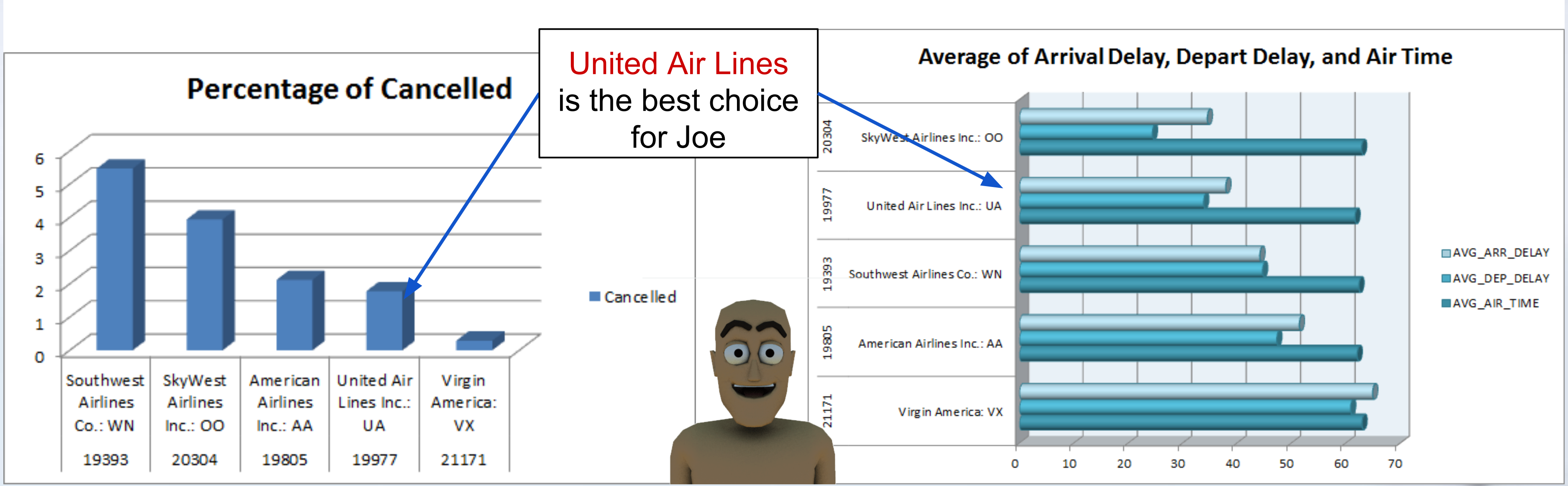
- San Francisco International Airport (Los Angeles)
- 5 airlines

- Joe calls his friend again
- Now, delay is ok, but Joe wants to take the airline with low average delay time and lowest cancelled rate





Los Angeles to San Francisco Contd…



United Air Lines is the best choice for Joe

References

- http://willowxdanimation.blogspot.com/2011/05/emotions-and-poses.html
- http://codename-animator.blogspot.com/
- http://www.free-power-point-templates.com/airline-powerpoint-template/

Thank You