GENERAL ASSEMBLY

FINAL PROJECT

Nicole Baez - Final Project Pt., Lightening Presentation
General Assembly - Dat2 | March 15th, 2016

REQUIREMENTS 2

- **The Problem:** What's the background and scope of the project idea? What problem are you attempting to address or solve? Who may it matter to?
- **Data:** What data exists to help solve this problem? Where is it coming from? What does the data look like? What is the observation?
- **Hypotheses:** Given the problem and data you're aware of, what do you believe is the solution? What does success look like?
- **Detailed Breakdown:** Explain three potential projects, including: a problem statement, a hypothesis, and potential data set and demonstrate familiarity with the domain of the data.

THE PROBLEM

- Delta receives approximately 30M unique visitors on their site per month from a variety of referral sources, campaigns and offline promotions. While many significant site actions are captured, I believe only a few actions during their visit that are reported actually help users book a flight or make an ancillary purchase.
- Segmenting and seeing reports of various customer journey for a variety of users is ideal, but filtering through the data and customer paths only suggests and hints on certain behaviors it does not predict, create relationships or help optimize conversion.
- **Goal: This project attempts to clarify customer journey points on-site that lead to the highest conversion based on their relationship to revenue. Stakeholders from various digital teams would find this useful.**

DATA

- This project will only use data exported from Omniture, which is click-stream data.
- Data is a .csv file with at least 2 years worth of daily activity.
- For the initial analysis daily data will include the following features: Unique visitors, revenue, average page depth, flight searches, fare change errors and the outcome variable is tied to daily revenue.

DATA DESCRIPTION


```
path = "./DeltaDS1.csv"
```

In [2]:
```
table = read_csv_file(path)
```

In [3]:
```
table.head()
```

Out[3]:

	Unique Visitors	Average Page Depth	Flight Searches	Fare Change Error	Revenue
0	695045	19	540686	514	22403850
1	747063	19	557525	537	22632654
2	1108106	14	873056	1034	37367126
3	1104659	15	917531	1259	42414291
4	1124734	15	895169	1349	41818597

In [5]:
```
print table.head()
```

```
   Unique Visitors  Average Page Depth  Flight Searches  Fare Change Error  \
0           695045                  19           540686                514
1           747063                  19           557525                537
2          1108106                  14           873056               1034
3          1104659                  15           917531               1259
4          1124734                  15           895169               1349

    Revenue
0  22403850
1  22632654
2  37367126
3  42414291
4  41818597
```

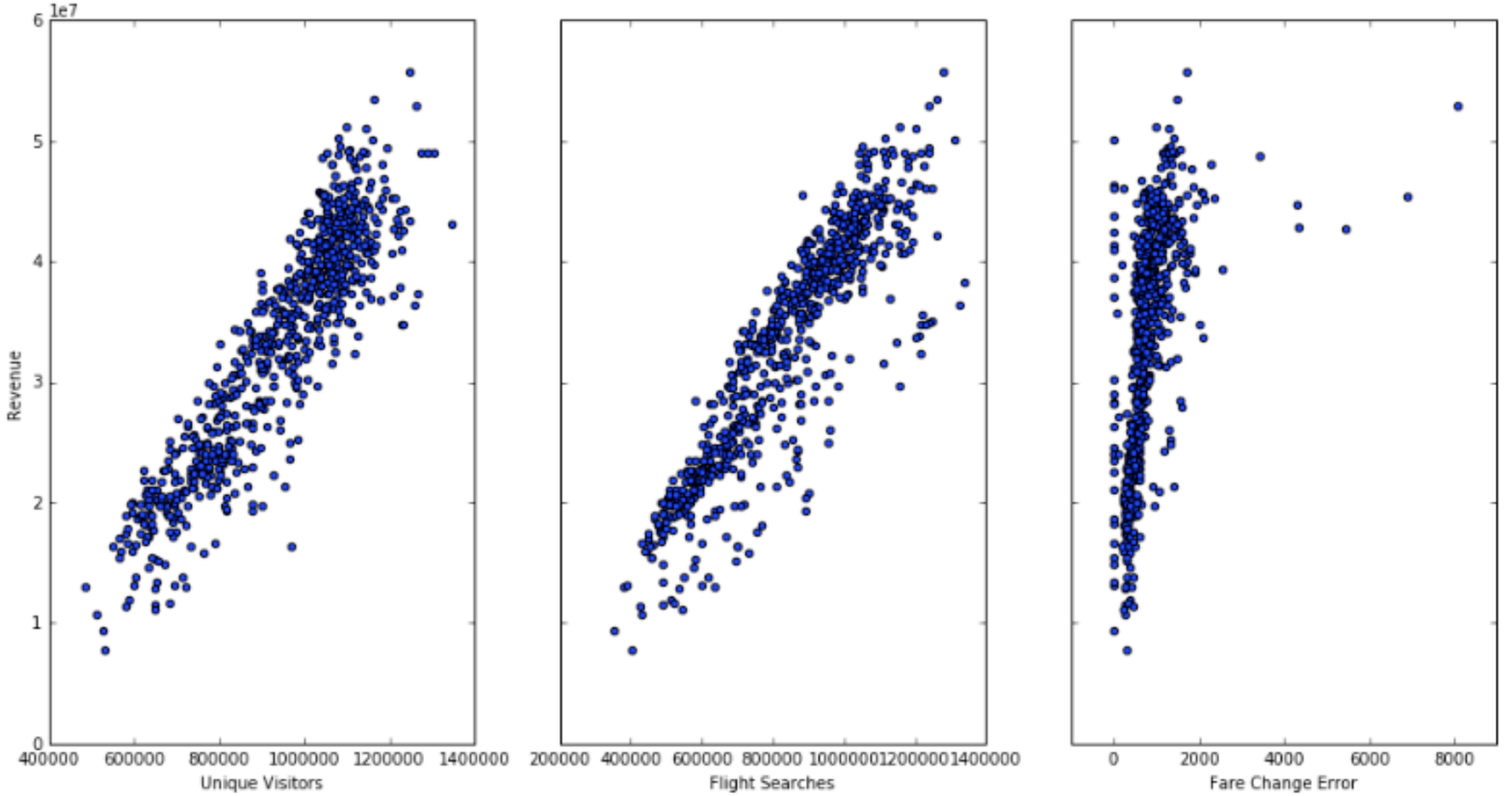
In [7]:
```
table.shape
```

Out[7]: (729, 5)

DATA VISUALIZATION

In [20]:
```
# visualize the relationship between the features and the response using scatterplots
fig, axs = plt.subplots(1, 3, sharey=True)
table.plot(kind='scatter', x='Unique Visitors', y='Revenue', ax=axs[0], figsize=(16, 8))
table.plot(kind='scatter', x='Flight Searches', y='Revenue', ax=axs[1], figsize=(16, 8))
table.plot(kind='scatter', x='Fare Change Error', y='Revenue', ax=axs[2], figsize=(16, 8))
```

Out[20]:
```
<matplotlib.axes._subplots.AxesSubplot at 0x11f188090>
```



HYPOTHESIS

- By using a few broad features in the data set I want to determine if there is any kind of relationship to revenue. The goal is to understand which step in the customer journey has the greatest impact on conversion. I believe traffic, site consumption, search activity or efficiency can all have an impact on conversion, but to varying degrees. Success for this project means a simple, clear understanding of which component of a visitors path impacts their conversion behavior at the most.

- **Feature1_Unique visitors:** Determine if the quality or amount of traffic that the site receives each day impacts conversion. Next steps could include further segmenting the traffic by source and flight type.
- **Feature2_Average page depth:** Are visitors who view more pages more likely to purchase? If so, what pages are they looking at most and why?
- **Feature3_Flight searches:** Does the amount of flight searches affect revenue? Which flight searches by destination or airport have the highest revenue.
- **Feature4_Fare change errors:** Does revenue change according to fare change errors decreasing or increasing?

DETAILS BREAKDOWN - 1

- **Feature1_Unique visitors**

Statement: Using daily unique visitor data from January 2014 - January 2016 determine the sources of traffic that drive the highest site conversion by flight type.

Data set: A variety of referral sources will include direct, paid, campaigns, internal and email. Flight types can include vacations, domestic and international travel.

DETAILS BREAKDOWN - 2

- **Feature2_Average page depth:**

Statement: Using daily page depth data from January 2014 - January 2016 determine the content, site sections and page visit frequency that yield the highest conversion before a purchase takes place.

Data set: Flight search destinations, days between flight search type and purchase, and site entry from flight promotion to flight searches on-site.

DETAILS BREAKDOWN - 3

- **Feature3_Flight searches**

Statement: Using daily flight search data from January 2014 - January 2016 determine the frequency and types of flight searches that yield the highest conversion before a purchase takes place.

Data set: Flight search destinations, days between flight search type and purchase, and site entry from flight promotion to flight searches on-site.

DETAILS BREAKDOWN - 4

- **Feature4_Fare change errors**

Statement: Using fare change errors data from January 2014 - January 2016 determine if errors experienced by site visitors increase or decrease their tendency to complete a purchase.

Data set: Total fare change errors, fare error types, flight error types.