

THE GEORGE  
WASHINGTON  
UNIVERSITY  
WASHINGTON, DC

# Company Sales Analysis

## Group 1

David Gayman  
Milad Nazar  
Ibrahima Diallo  
Jonathan Feldmann  
Jacinta Oduor



# Outline

1. Project Introduction
2. Data & Resources Gathering
3. Data Cleanup & Exploration
4. Data Analysis
5. Discussions & Key Conclusions
6. Q&A

# Project Introduction

## *Motivation*

- Opportunity to provide data analysis for a sales scenario.

## *Hypothesis*

- A thorough data analysis has the ability to identify factors associated with high levels of sales and conversions, as well as provide projections of future income.

# Project Introduction

## *Company Profile*

- Provides a subscription-based service.
- Provides subscription terms ranging from one month to three years.
- Has recently experienced an uptick in sales due to pandemic-related factors.

# Project Introduction

## *Key Questions*

- (a) "What trends and patterns can be identified from the data?"
- (b) "What aspects of compliance and due diligence are in place to protect disruption of business?"
- (c) "Is the company targeting the highest-responding customers?"
- (d) "What unforeseeable factors are likely to affect future sales?"
- (e) "What market penetration does the company have into potential customers?"
- (f) "What market penetration does the company have regionally?"
- (g) "How are the highest paying and highest responding customers identified and classified?"
- (h) "What lessons have been learned during the analysis?"
- (i) "What recommendations are suggested to ensure increased future sales?"
- (j) "What is a likely projected range of future sales?"

# Project Introduction

## *Deliverables*

- Project Proposal & Work Plan
- ReadMe File
- Presentation
- Dashboard

Category	Goal #	Goal Name	Goal	Task #	Task Priority	Task Name	Assignee
Data Ingestion	G010	Required raw data files collected > 3000 repository.		1010	11/02/2020	Configure repository.	
				1020	11/02/2020	Document raw data files with description of content, as well as sourcing.	
				1030	11/02/2020	Get sales data into raw data directory.	
				1040	11/02/2020	Get marketing communications/campaigns/responses into raw data directory.	
				1050	11/02/2020	Get list of schools that use customer into raw data directory.	
				1060	11/02/2020	Get list of all schools that use possible customers into raw data directory.	
				1070	11/02/2020	Get school locations into raw data directory.	
				1080	11/02/2020	Read raw files into DataFormat.	
				1090	11/02/2020	Perform basic cleanup (null, invalid characters, etc.).	
				1100	11/02/2020	Perform advanced cleanup (collected/completed/processed).	
Data Processing	G020	Infrastructure implemented.		2010	11/02/2020	Implement infrastructure to provide data access to all developers.	David
				2020	11/02/2020	Implement infrastructure to provide collection and handling of processing artifacts (processed data, metrics, and plots -> multiple dashboards, compiled reports).	David
				2030	11/02/2020	Use one new Python library that hasn't been covered in class.	
				2040	11/02/2020	Use one new Python library that hasn't been covered in class.	
				2050	11/02/2020	Use one new Python library that hasn't been covered in class.	
				2060	11/02/2020	Use one new Python library that hasn't been covered in class.	
				2070	11/02/2020	Use one new Python library that hasn't been covered in class.	
				2080	11/02/2020	Use one new Python library that hasn't been covered in class.	
				2090	11/02/2020	Use one new Python library that hasn't been covered in class.	
				2100	11/02/2020	Use one new Python library that hasn't been covered in class.	

# Project Introduction

## *Solution Approach*

1. Identified key questions and prepared a Work Plan.
2. Obtained data from the customer.
3. Accomplished work through team discussions, individual contributions, pair programming.

# Data & Resources Gathering

- Raw data sourced and collected into Resources/Raw repository directory.
- Raw files exported to .csv in the Raw directory.



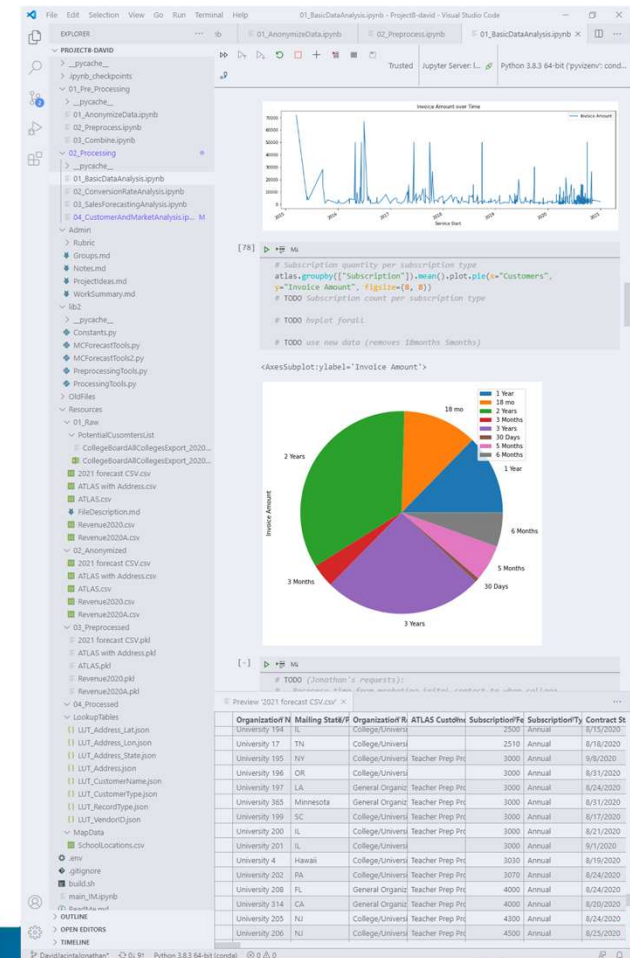
# Data Cleanup & Exploration

## Technologies

Python was used inside Jupyter Lab notebooks to provide dynamic data analysis.

## File Organization

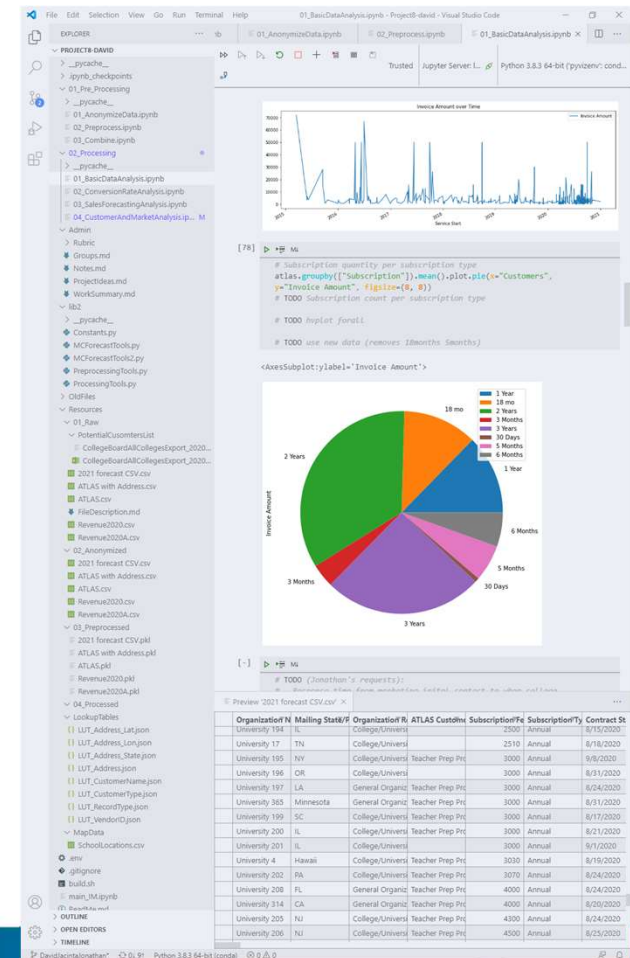
- Data files placed in Resources.
- Jupyter notebooks organized by processing stage.
- Data shared across notebook collection via binary files containing serialized (pickled) DataFrames.
- Tools implemented in shared Python class files.



# Data Cleanup & Exploration

## Processing Stages

1. Raw Data: Raw unprocessed data.
2. Anonymized: Raw data with customer names anonymized.
3. Pre-processed: Preprocessed, cleaned data.
4. Processed & LookupTables: Processed data including lookup tables generated from preprocessed data to enhance data merging.

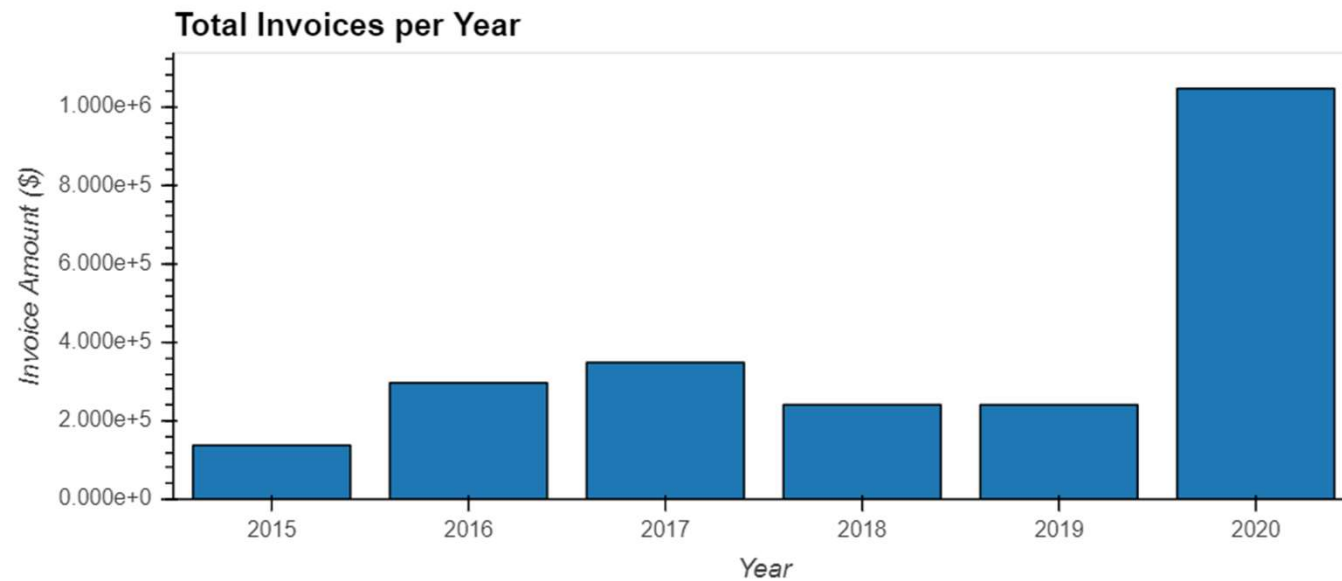


# Data Analysis: General Analysis

(a) *“What trends and patterns can be identified from the data?”*

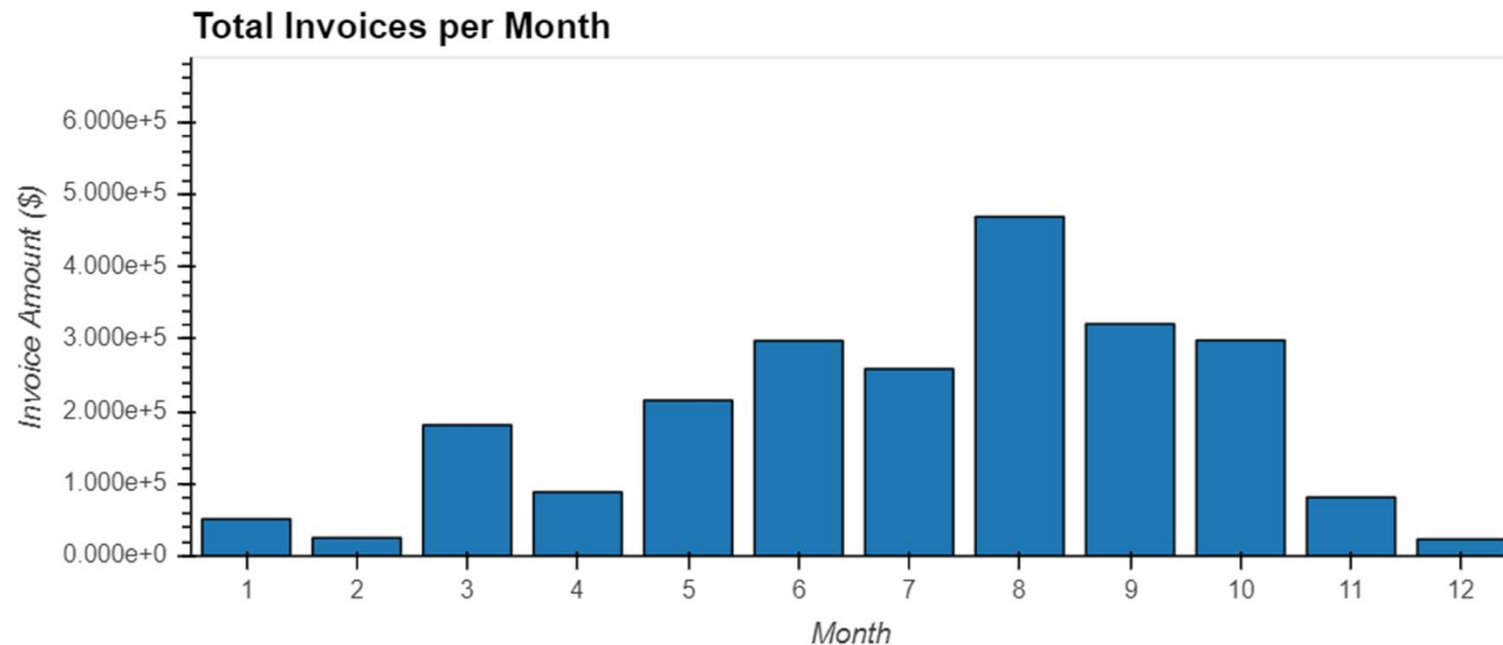
# Data Analysis: General Analysis

Sales Increase in 2020



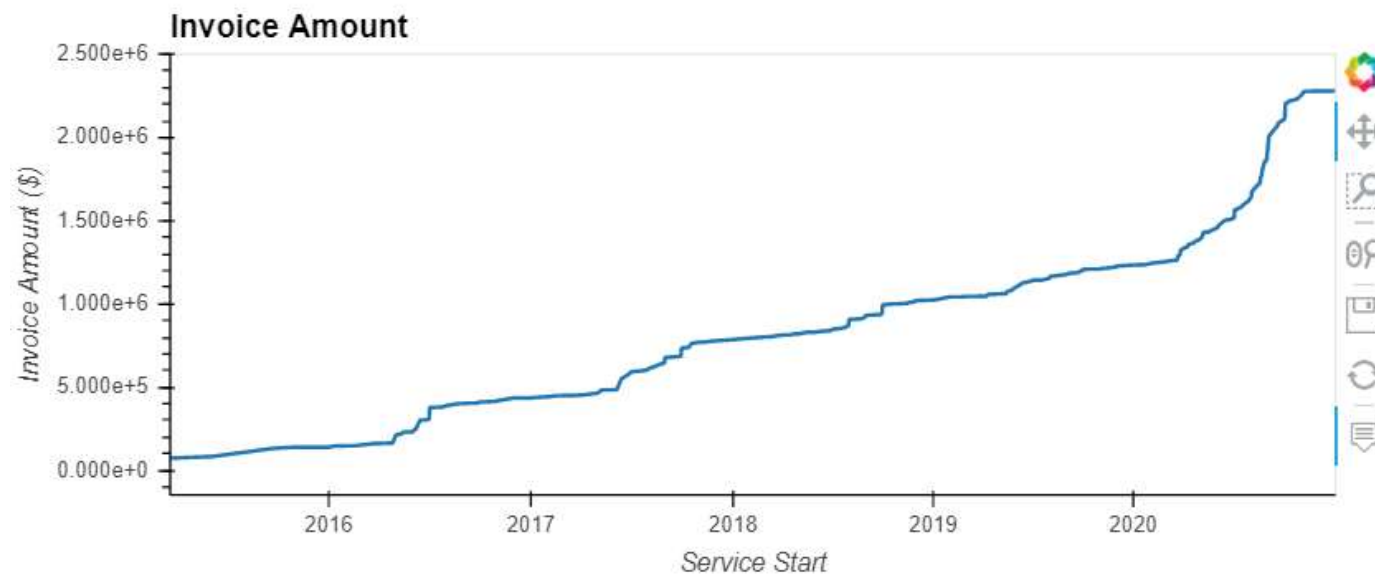
# Data Analysis: General Analysis

Best-Selling Months (2015 – 2020)



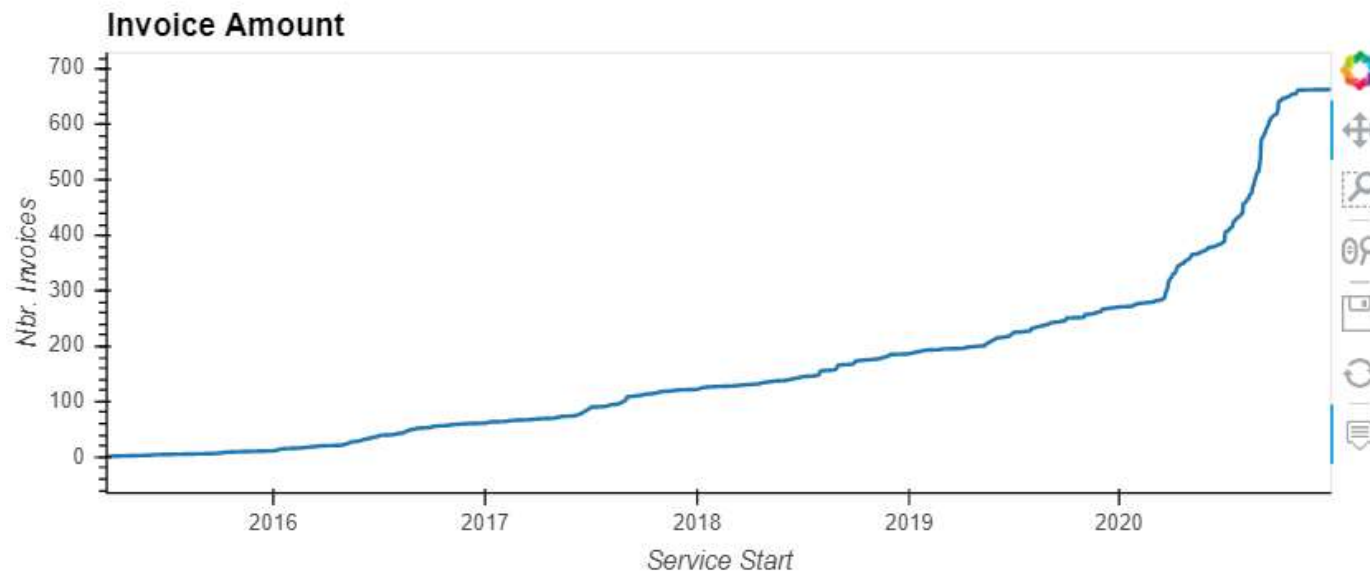
# Data Analysis: General Analysis

Aggregate Sales (Invoice Amount, \$) 2015 – 2020



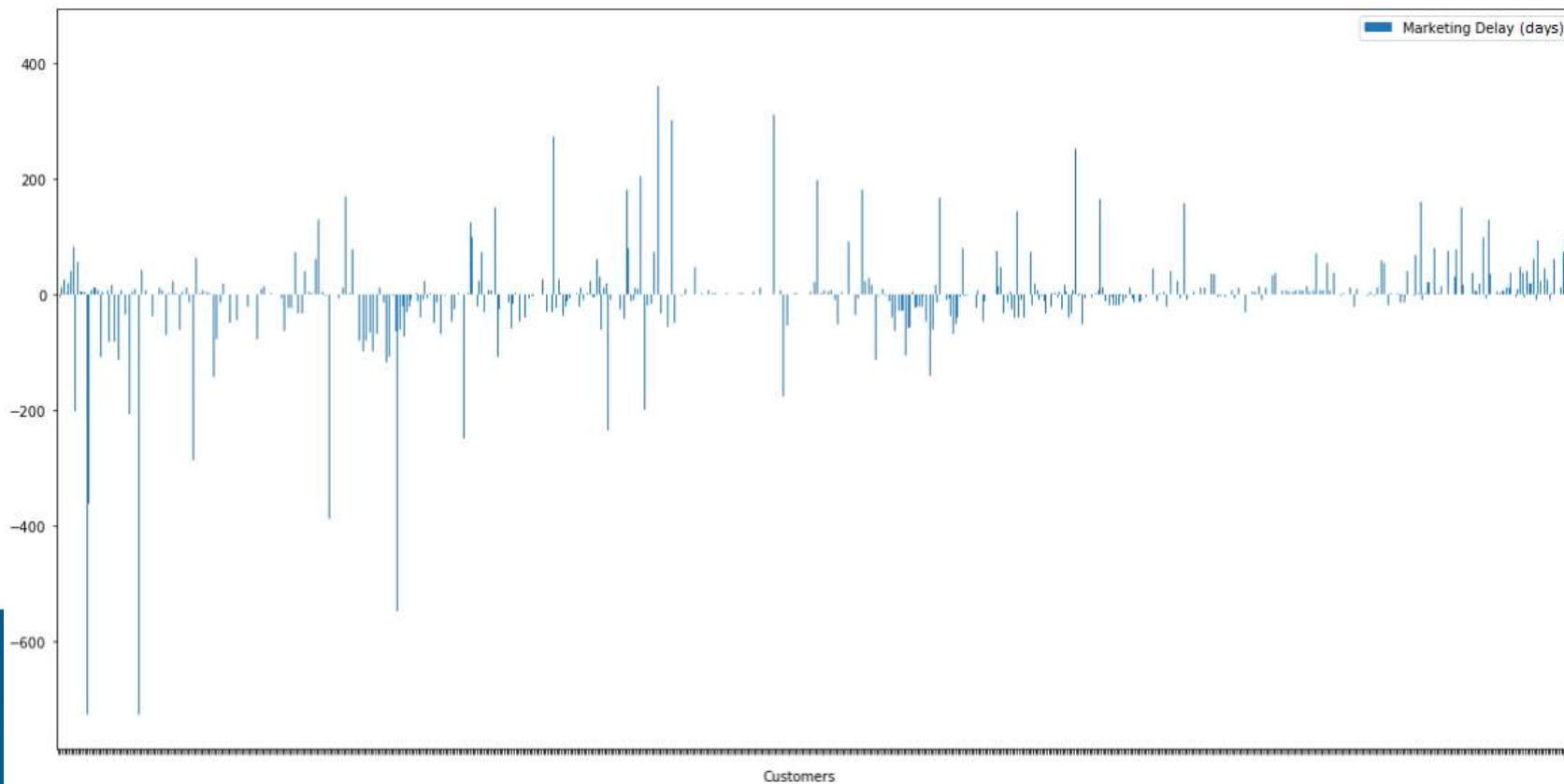
# Data Analysis: General Analysis

Sales (Number of Invoices) 2015 – 2020



# Data Analysis: Conversion Rate Analysis

(c) *“Is the company targeting the highest-responding customers?”*

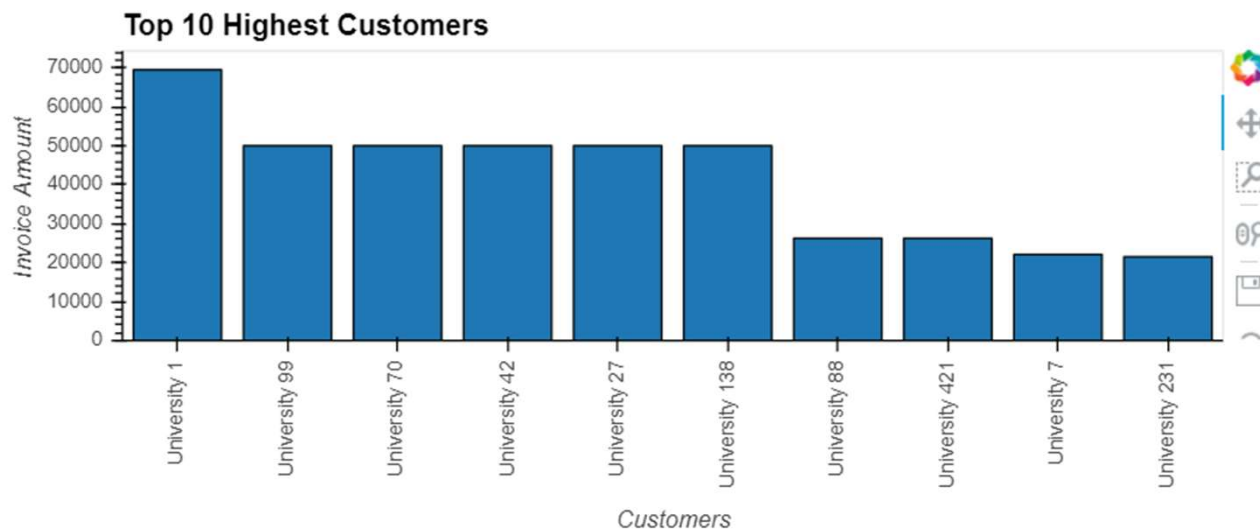




# Data Analysis: Conversion Rate Analysis

(c) “Is the company targeting the highest-responding customers?”

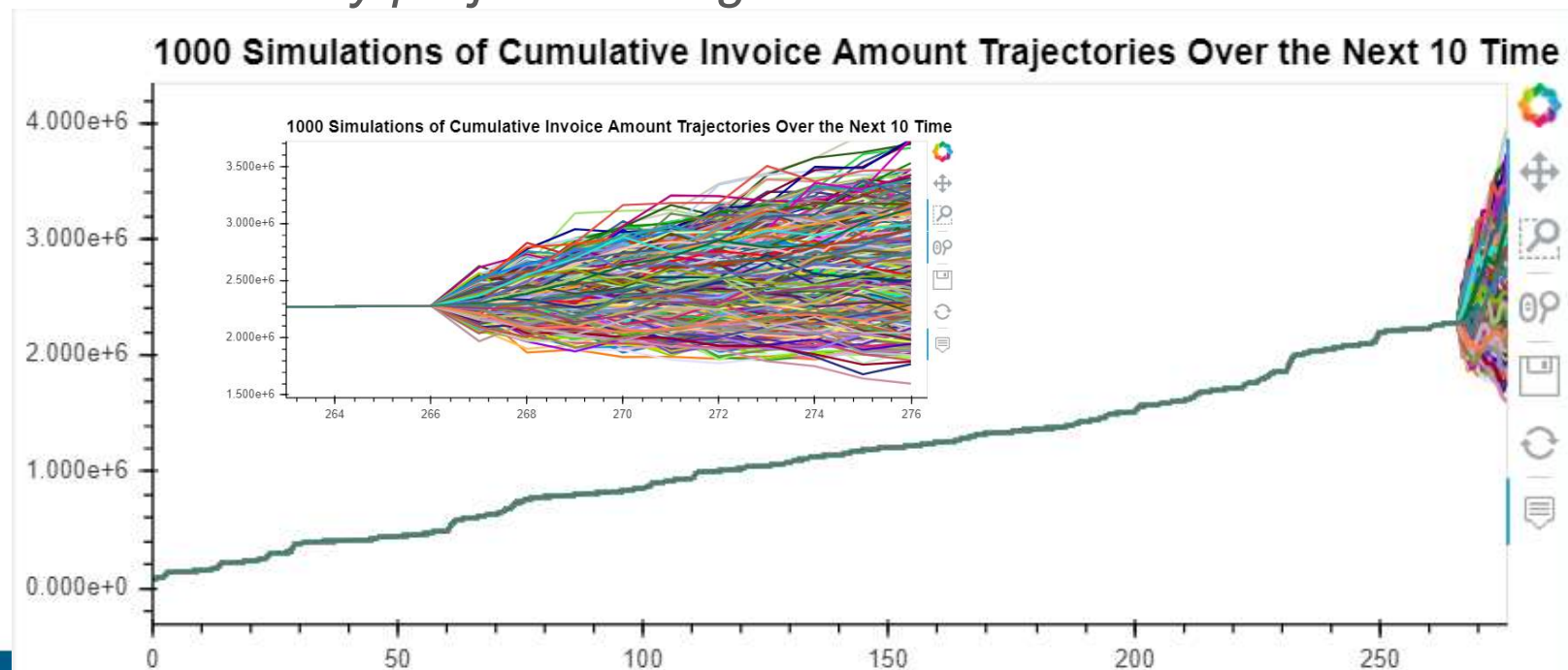
(g) “How are the highest paying and highest responding customers identified and classified?”



***Top 10 Paying Customers(2015 – 2020)***

# Data Analysis: Sales Forecasting Analysis

(j) *“What is a likely projected range of future sales?”*



# Data Analysis: Sales Forecasting Analysis

(j) *“What is a likely projected range of future sales?”*

- Sales projections computed using a legacy method are available for comparison to the Monte Carlo simulation. (proprietary)
- Sales Potential: Future earnings potential may or may not be correlated with historical trends.

# Data Analysis: Customer & Market Analysis

(e) *“What market penetration does the company have into potential customers?”*

(f) *“What market penetration does the company have regionally?”*

# Data Analysis: Customer & Market Analysis

- Total market capture computed based on current vs. potential customers:

Total Market Capture Report:

The number of paying customers is 719

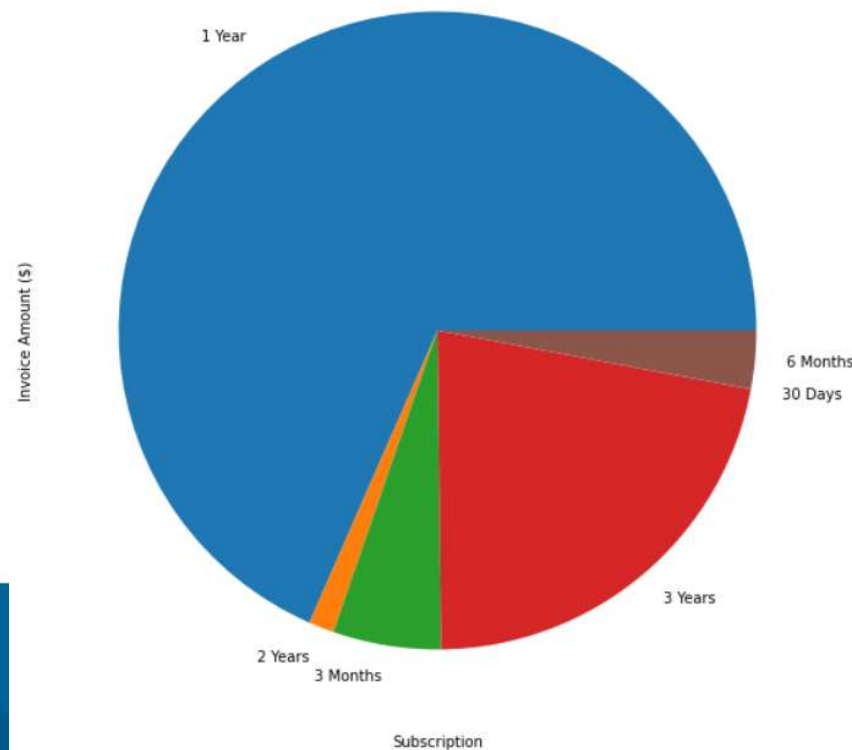
The number of potential customers is 3835

Market penetration is 18.748370273794002 %

# Data Analysis: Customer & Market Analysis

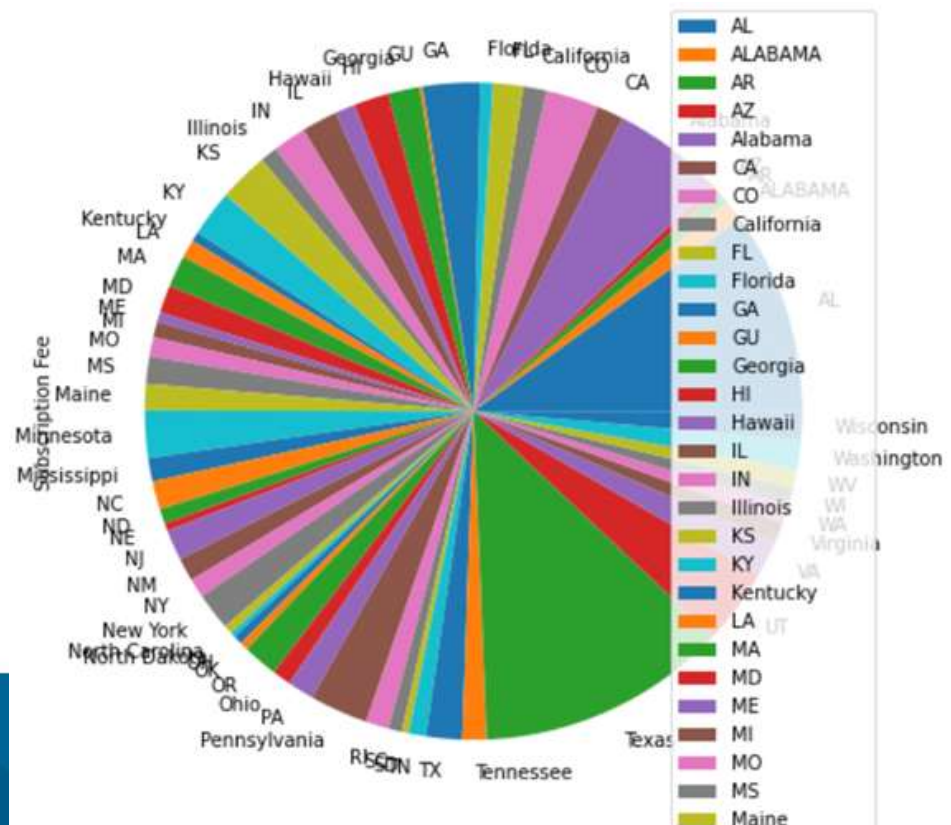
## Subscription Earnings Breakdown

Subscription Income (\$) from 2015 to 2020



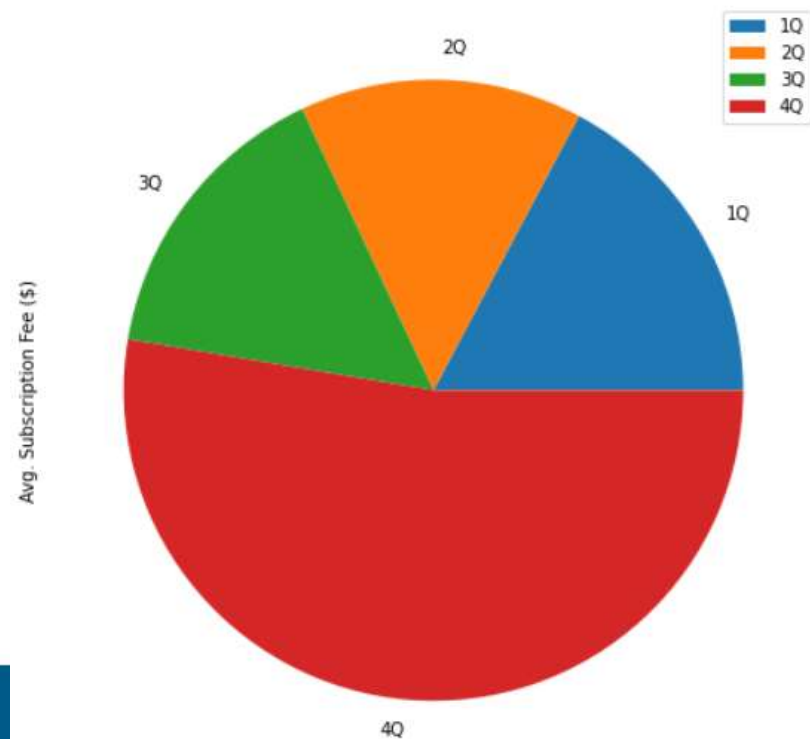
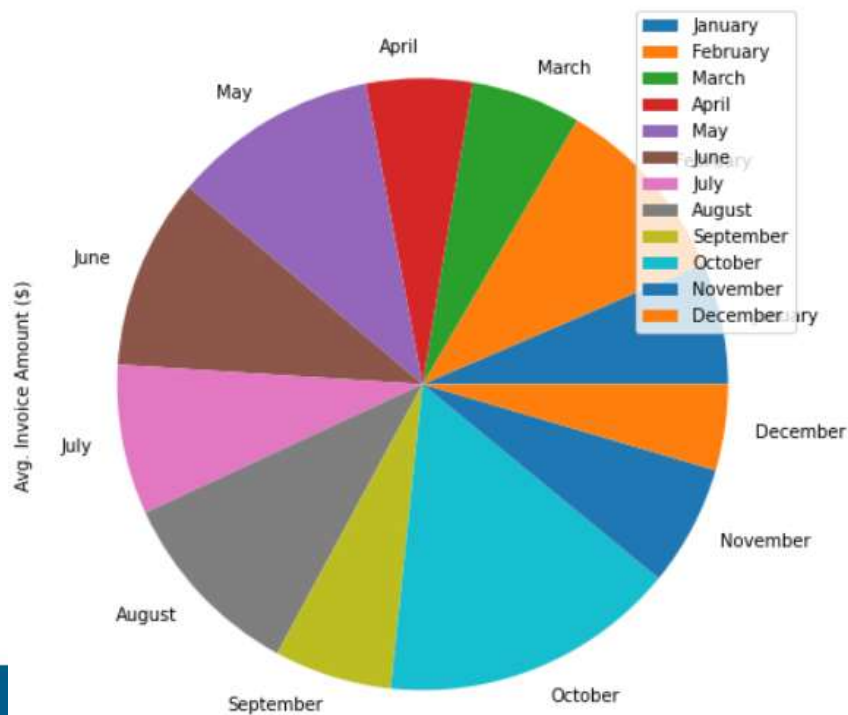
# Data Analysis: Customer & Market Analysis

## Customer Grouping by State



# Data Analysis: Customer & Market Analysis

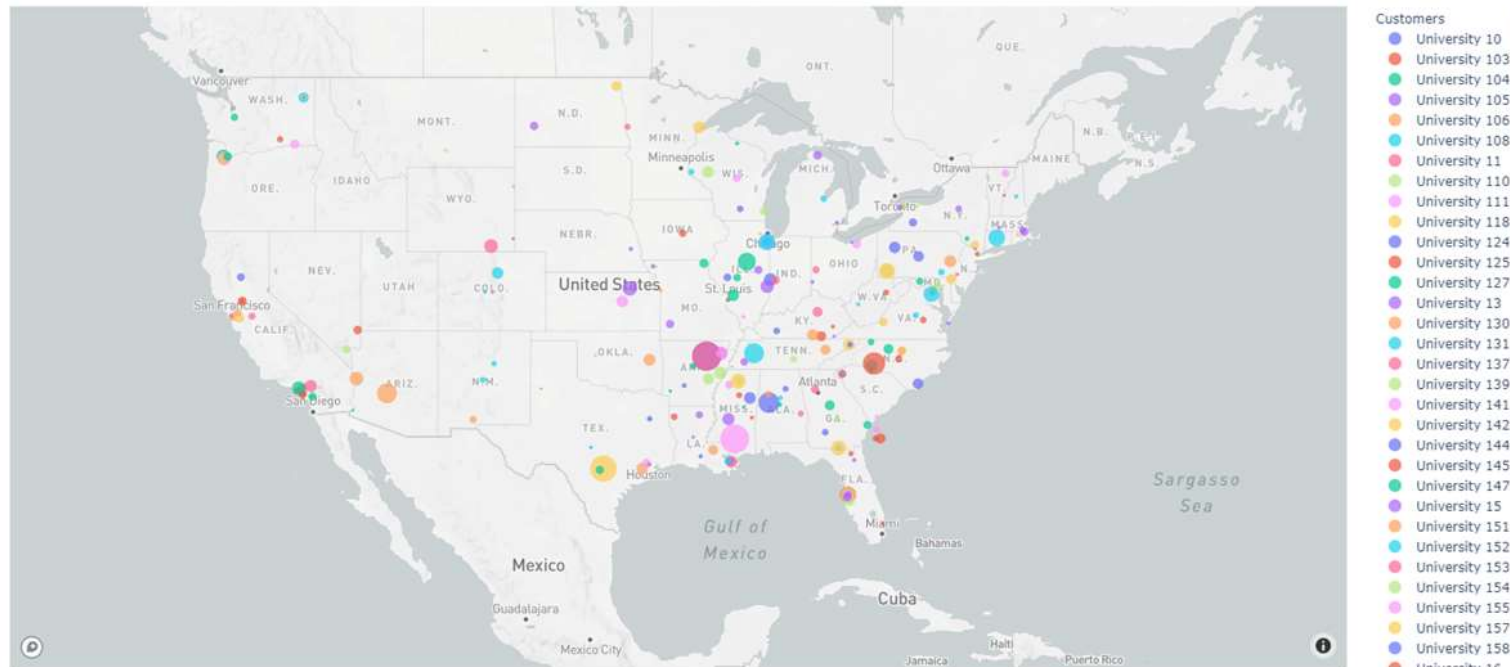
## Subscription Type Breakdown





# Data Analysis: Customer & Market Analysis

Geographic Customer Capture – U.S.



# Data Analysis: Customer & Market Analysis

## Geographic Customer Capture - World



# Discussions & Key Conclusions

*(h) “What lessons have been learned during the analysis?”*

- A clear increase in sales occurred in 2020, which may be associated with pandemic-related factors. Moving forward, special care should be taken to:
  - Look for factors that led to the highest-responding customers, and focus on further expansion.
  - Even with the massive 2020 expansion in sales, much of the customer base remains untapped; the recommendation is to market to potential customers that are unaware of the product.
- The current product offering contains limited complexity and associated products; look for additional related products and services that can be marketed for add-on sales.

# Discussions & Key Conclusions

*(i) “What recommendations are suggested to ensure increased future sales?”*

- Target marketing communications to customers who have a historically higher conversion rate, number and value of purchases.
- Current subscriptions automatically expire; consider automatically re-purchasing subscriptions at expiration.

# Questions?

## Q & A