David Gayman

davidrgayman@gmail.com

ATLAS Analysis

Project Plan

# Overview

TODO rename to project proposal

TODO make sections:

* Objective (<https://github.com/coding-boot-camp/GWU-ARL-FIN-PT-09-2020-U-C/blob/master/08-Project-1/Project-01/ProjectGuidelines.md>)
  + Problems to solve:
    - Sales projection
    - Correlation for sales data
    - Sales coverage map / insights
  + Questions to answer:
* Process/Methodology:
  + Using data provided by the client, data analysis will be performed (leveraging class examples).
  + As part of the analysis, do a review/walkthrough of existing class examples.
* Expected Outcomes:
  + 6-8 analytics (plots, data etc.),
  + analysis & conclusions;
  + presentation
* Tasks & assignments
* Timeline (1.5 weeks)

# Goals & Tasks

TODO

* Review:
  + Tasks from paper notes
  + Tasks from sticky notes
  + Tasks from repository files
  + Project instructions

## Timeline

|  |  |  |
| --- | --- | --- |
| Stage | Name | Description |
| I | Setup | Implement the basic data import and infrastructure. |
| II | Analysis | Multiple Goals can be worked on at the same time by different people. |
| III | Conclusions | Draw conclusions and merge into final report and dashboard. |

## Key

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| |  |  | | --- | --- | | Priority | Meaning | | 1 | High | | 2 | Medium | | 3 | Low |   **Priority** | |  | | --- | | Normal item | | Complete | | Extremely important item | | Critical-path item |   **Color** |

## Goal & Task Data

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Category | Goal # | Goal Priority | Goal | Task # | Task Priority | Deadline | Task | Assignee |
| Data Ingestion | G010 |  | Required raw data files collected in repository. | T010 |  | 11/5/2020 | Configure repository. |  |
|  |  |  | Document raw data files with description of content, as well as sourcing. |  |
| T020 |  | 11/5/2020 | Get sales data into raw data directory. |  |
| T030 |  | 11/5/2020 | Get marketing communications/attempts/responses into raw data directory. |  |
| T040 |  | 11/5/2020 | Get lists of schools that are customers into raw data directory. |  |
| T050 |  | 11/5/2020 | Get List of all schools that are possible customers into raw data directory. |  |
| T060 |  | 11/5/2020 | Get school locations into raw data directory. |  |
| G020 |  | Data ingestion implemented. | T070 |  | 11/5/2020 | Read raw files into DataFrames. |  |
| T080 |  | 11/7/2020 | Perform basic cleanup (nulls, invalid characters, etc.). |  |
| T090 |  |  | Perform advanced cleanup (collect/combine/pre-process). |  |
| Data Processing | G030 |  | Infrastructure implemented. | T100 | 1 | 11/7/2020 | Implement infrastructure to provide data access to all developers. | David |
| T110 | 1 |  | Implement infrastructure to provide collection and handling of processing artifacts (processed data, metrics, and plots -> multiple dashboards, compiled reports). | David |
|  |  |  |  |  |  |  | Use one new Python library that hasn't been covered in class (math) (import from excel files!!!). |  |
|  |  |  |  |  |  |  | Optionally, use at least one API, if you can find an API with data pertinent to your primary research questions. |  |
|  |  |  |  |  |  |  | Implelent database storage. |  |
| Basic Data Analysis | G040 |  | Analysis procedure and plots implemented. | T120 | 1 | 11/10/2020 | Implement analysis procedure and plots. |  |
| T130 | 2 |  | Plot trends showing comparison to other educational services. () |  |
| G050 |  | Conclusions documented in the final report. | T140 | 1 |  | Answer in the report: “Which quarters, years etc. have the highest sales, and highest growth?” |  |
| T150 | 2 |  | Answer in the report: "What conclusions can be reached by analyzing the trends in comparison to other educational services?" |  |
| T160 | 1 |  | Answer in the report: "What actions can be taken to increase sales?   * Suggestions: Send more/better targeted emails, phone calls, target certain types of customers, would this be applicable to ... * Suggest 2-3 actions that can be taken to increase sales, and justify with plots. |  |
| Conversion Rate Analysis | G060 |  | Analysis implemented. | T170 |  |  | Document the sales process (initial contact, follow-up, negotiation, purchase). () |  |
| T180 |  |  | Compute time differences within stages of the sales process. () |  |
| T190 |  |  | Compute the initial purchase conversion rates. () |  |
| T200 |  |  | Compute the upgrade conversion rates. () |  |
| G070 |  | Conclusions documented in the final report. | T210 | 1 |  | Answer in the report: "What is the length of time between initial contact and purchase decision?" | Ibrahima & David |
| T220 | 1 |  | Answer in the report: "What is the initial purchase conversion rate based on marketing/sales communications?" | Milad & Jonathan |
| T230 | 1 |  | Answer in the report: "What is the upgrade conversion rate (from 3-months to 6-months to 1-year)?" | Milad & Jonathan |
| Sales Forecasting Analysis | G080 |  | Analysis implemented. | T235 |  |  | Identify parameters to vary as part of Monte-Carlo analysis. (T260) |  |
| T240 |  |  | Implement Monte-carlo-based sales forecasting procedure and plots. (T260) | Milad & Jonathan |
| T250 |  |  | Implement legacy sales forecasting procedure and plots. (T260) | Milad & Jacinta |
| G090 |  | Conclusions documented in the final report. | T260 | 1 |  | Answer in the report: "Taking into account both legacy and Monte-Carlo analyses, what sales are predicted for each quarter through end of 2021?" | Milad & Jonathan |
| Customer & Market Analysis | G100 |  | Analysis implemented. | T270 |  |  | Categorize school data (hint: DataFrame.groupby()). () |  |
| T275 | 1 |  | Verify if mapbox API handles city/addresses; if not, identify customer lat/lon location. |  |
| T280 | 1 |  | Plot customers on a map including interactive selector for the year. |  |
| T290 |  |  | Analyze data to identify highest-responding, and highest-paying customers. |  |
| T300 |  |  | Correlate conversion rates with the customers ranked by highest-responding/highest-paying. |  |
| G110 |  | Document conclusions in the final report. | T310 | 2 |  | Answer in the report: "What market penetration does the company have into each group of schools (ivy league, state schools, community college, high schools, etc.)" |  |
| T320 | 1 |  | Answer in the report: "What market penetration does the company have regionally?" |  |
| T330 | 3 |  | Answer in the report: "Is the company targeting the highest-responding customers?" |  |
| T335 |  |  | “What aspects of compliance and due dilligence are recommended?” | Jacinta |
| Deliverables | G120 |  | Final Report | T340 |  |  | Identify and describe the conclusions of the analysis. | All |
| T350 |  |  | Describe the specific analysis procedure used to support the analysis. | All |
| T360 |  |  | Document the specific data sources, resources, links, etc. that were used for analysis, in the appendix. | All |
| T370 |  | 11/16/2020 | Provide report introduction, table of contents, conclusion, and appendix in the deliverable zip file. | Milad |
|  |  | ReadMe File |  |  |  | Write a project summary that looks nice with pictures. Create a README.md in your repo with a write-up summarizing your major findings. This should include a heading for each question you asked of your data and under each heading a short description of what you found and any relevant plots. |  |
|  |  | Project Proposal |  |  |  |  |  |
|  |  | Presentation |  |  |  | Title Slide | Jacinta |
|  |  |  |  |  | Motivation & Summary Slide | Jacinta |
|  |  |  |  |  | Questions & Data | Jacinta |
|  |  |  |  |  | Data Cleanup & Exploration | Jacinta |
|  |  |  |  |  | Data Analysis | Jacinta |
|  |  |  |  |  | Discussion | Jacinta |
|  |  |  |  |  | Postmortem | Jacinta |
|  |  |  |  |  | Questions | Jacinta |
|  |  |  |  |  | Additional presentation requirements (https://github.com/coding-boot-camp/GWU-ARL-FIN-PT-09-2020-U-C/blob/master/08-Project-1/Project-01/PresentationRequirements.md) |  |
|  |  |  |  |  | Deliver the presentation | Milad & David |
| G130 |  | Dashboard | T380 | 1 |  | Implement a dashboard structure to organize the analysis results and data views. |  |
| T390 | 1 |  | Implement analsysis-specific dashboard views containing analysis plots. | All |
| T400 | 2 |  | Implement college-specific dashboard view(s) that display analysis plots for the college in the selected drop-down. |  |
| T410 | 3 |  | Implement college-specific dashboard view(s) that display pre-processed (unanalyzed) data plots for the college in the selected drop-down. |  |
| T420 | 2 |  | Implement served/hosted dashboard. |  |
| T430 | 2 |  | Document how to access the hosted dashboard (if applicable). |  |
| T440 | 1 |  | Package the dashboard solution in the deliverable zip file. |  |
| G140 |  | Source code | T450 | 2 |  | Package the source code in the deliverable zip file. |  |
| G150 |  | Processing artifacts | T460 | 3 |  | Implement data dumps for all critical plots. |  |
| T470 | 2 |  | Package the important processing artifacts in the deliverable zip file. Examples include key metrics, statistics, data dumps, etc. |  |