CASE STUDY: DELTAPOLL

Implementing a Modern Data Pipeline - From Raw Data to Actionable Results

AT A GLANCE

MAIN GOALS

Help the client create a platform for growth that removes tedious tasks and lets them focus on providing insight.

METHOD

Use Datasmoothie's APIs and platform to digest data files and automate the processing, reporting and publishing pipeline.

RESULT

A rich, automatable, Python based data pipeline creating

THE CLIENT

Deltapoll is a rapidly expanding independent public opinion consultancy founded in 2017. They focus on using cutting edge methods and new technology coupled with unrivalled experience, to provide bespoke research, analysis and strategic counsel tailored to the individual needs of their clients.

OPPORTUNITY

As an expanding company, Deltapoll was facing several challenges scaling their existing data workflow to their growing requirements:

Onboarding: SPSS had proven to be an invaluable tool in the creation and initial growth of the company. Nevertheless, with the hiring of new staff, a desktop based tool was proving difficult to scale across new resources. Getting new staff members up to speed would mean installation and set-up on each new device.

Training and Support: SPSS was once the dominant tool in data analysis but has long been supplanted by more modern methods. The dwindling number of SPSS users has been matched by dwindling peer support and wider resources.

Cost and Features: SPSS operates on a model that necessitated buying additional modules in order to access full functionality for tabular data analysis. With a growing team, these additional costs were adding up.

datasmoothie

client-ready deliverables without the tedious manual formatting.

ABOUT US

Datasmoothie is an award-winning platform to crunch and visualise data. Datasmoothie was founded in 2015 and has won the MRS award for technical innovation. Datasmoothie has worked with some of the world's largest market research companies on data crunching and visualisations.

PRICING MODEL

Datasmoothie is both a SaaS platform and agency partner with pricing models that reflect both aspects.

ABOUT US

Datasmoothie is blended in London and Reykjavík.

Get in touch at:

info@datasmoothie.com

Flexibility: The current Deltapoll data pipeline requires that they use a data collection agency that can deliver an SPSS file to them.

Repetition and Automation: The Deltapoll data pipeline can be broken down into a series of steps: cleaning, recoding, weighting, analysis, and reporting. SPSS allowed little room for the automation of repetitive tasks.

Secure online environment with source control and automation.

Deliverables: Whilst SPSS enabled Deltapoll to execute their data pipeline, the endpoint of the SPSS workflow was not a client-ready deliverable. Staff members would have the tedious task of turning the SPSS output into an Excel Table that was ready for the client to consume.

SOLUTION

Datasmoothie engaged with Deltapoll to fully understand their requirements and pain-points. As a result, Datasmoothie implemented a secure, hosted solution based on its Python Tally technology (this involved Datasmoothie taking all of Deltapoll's existing SPSS syntax files (.sps) and translating them into Python). Using Microsoft's GitHub Codespaces, Datasmoothie implemented a modern data pipeline with each step in the Deltapoll workflow transferred to its own Jupyter Notebook Document (A script file that implements and executes Python code in blocks, or cells). This pipeline can be run manually or through events-based (such as the downloading of new data) automation. After each Notebook has been run a series of automated tests can be instigated to check the quality of the pipeline and report any errors (Are bases too low? Are weight factors within thresholds? etc.).

Onboarding: Because the solution is based on Microsoft's GitHub Codespaces, whenever a new user joins they just need

to sign in with their credentials and they are met with a ready-to-go, configured and templated environment. No time is wasted with installation and setup.

Training and Support: Datasmoothie provides extensive online documentation and support. Nevertheless, because Tally is a Python based solution there are a myriad of online training, support and reference materials to draw from meaning that if you have a problem, there is almost certainly someone with a solution to it.

Cost and Features: Tally has only one, up front, licence cost which gives you access to all of its features. In addition, Tally can be combined with other Python packages, to extend and enhance its core functionality.
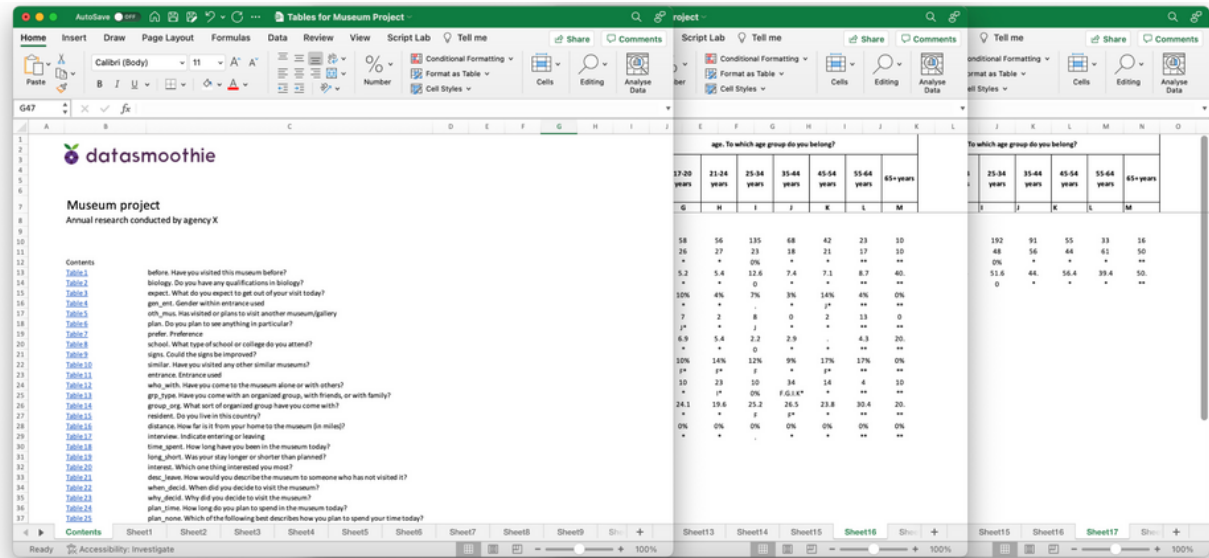
Flexibility: Tally is data source agnostic. You can use any of the common data formats. In fact, you can import data in various formats and analyse in a uniform manner just by adding one line of code. This has given Deltapoll the flexibility to switch data collection suppliers without impacting their workflow.

```python
# parameters
spss_file = "../data/Omni 102 week 1 Feb 2023 Final data.sav"
```
[3]   ✕ parameters   ＋ Tag

```python
ds.use_spss(spss_file)
ds.use_confirmit()
ds.use_csv()
ds.use_nebu()
ds.use_parquet()
ds.use_quantipy()
ds.use_unicom()
```
[4]

Repetition and Automation: Datasmoothie broke each phase of the Deltapoll data pipeline down into an individual Notebook: Cleaning, recoding, weighting, analysis and reporting were turned into automation phases so that analysts could focus on process exceptions and insight, rather than operational minutia.

Deliverables: Tally's rich API allowed Deltapoll to have fine-grained control over the appearance of their Excel

outputs, encoding the 'Deltapoll look and feel' in code, ready for automated execution, saving significant time in manual formatting.

Rich, engaging Excel client deliverables.

RESULTS & BENEFITS

The implementation of Tally helped Deltapoll to overcome their challenges and achieve the following results:

Fast, Secure, Scalable Onboarding : Users can be brought on to ready made and inherently secure workspaces. All user activity can be audited making it easy to demonstrate data pipeline integrity to clients.

Faster Data Processing: The software allowed the agency to process data much faster, reducing the time and effort required for manual data processing.

Flexible procurement: Being able to use different data formats with minimal code change has given Deltapoll the flexibility to switch data collection suppliers without impacting their workflow.

Increased Accuracy and Quality : by enabling Deltapoll to catch errors, anomalies and exceptions through automated testing,

Datasmoothie was able to speed up the time to discover and correct any issues with the data.

Enhanced Data Analysis Capabilities: Deltapoll has offered its clients enhanced data analytics by using R. However, the esoteric nature of SPSS meant that this capability had sat outside of their normal workflow. As a Python based solution, Tally offers complete interoperability with R, allowing the agency to quickly generate meaningful insights within their new standard workflow.

Improved Client-Ready Deliverables: One of the biggest pain-points for Deltapoll was the time it was taking to format outputs. Not only was this process time consuming for the researchers, it was also a tedious waste of skilled resources. By encoding the 'Deltapoll look and feel' in script, Datasmoothie was able to free the researchers to concentrate on delivering insight.

CONCLUSION

The implementation of Tally is a game-changer for Deltapoll. The software is helping the agency to overcome the challenges posed by their success and achieve significant improvements in their data processing and analysis workflow. Moreover, Deltapoll now has a platform for scalable growth being able to onboard extra resources whilst minimising associated operational costs at the same time as allowing those resources to concentrate on delivering insight.