Data-Driven and User-Centric Insights Function:

A Case of Statistics Centre - Abu Dhabi

Author: Abdulla Al Dhaheri

Co-author: H.E. Badria Hablail

Executive Summary:

The field of Statistics and Data Science has experienced a profound transformation as a result of breakthroughs in Data Science and Big Data. Data Science has revolutionized Data Analysis, offering swifter and more reliable insights from complex Data Sets. Big Data complemented and enhanced the revolution, enabling us to collect and analyze massive data quantities that was previously challenging to tap into. This paper seeks to elucidate the influence of these technologies on conventional governmental operations, spotlighting how entities like ours has leveraged Data Science tools to thrive in the contemporary data-driven world.

Aligned with Abu Dhabi's governmental strategy to transform Data into strategic assets, the Statistics Center - Abu Dhabi (SCAD) has developed the Insights and Foresights Platform (IFP). This platform aids decision-makers by equipping them with state-of-the-art analytical tools for analytical diagnosis, and visual demonstrations of certain trends, scenarios, and key official indicators. Our platform ingests Data from diverse sources, boasting advanced algorithms that enable data collection, thus enabling the dynamic integration of new a data spectrum. Concurrently, our expert team constantly collaborates with high-caliber synergies in the field of Data Science and Statistics to synchronize and enhance our analytical practices. By enhancing timeliness, accuracy, and our data range, as well as unveiling newly innovated methods to traditional statistical practice, we have collectively advanced in the frontier of innovation as a National Statistical Office. This has also empowered us to maintain our role in supporting our leadership as "the single source of trust".

The IFP holds immense potential for diverse sectors, including the private industry, academia, and the general populace. Users can gain a deeper understanding of market trends and consumer behavior, paving the way for a more collaborative practice between synergies. This will ultimately serve as a bridge of transparency and effective collaboration in our Statistical Ecosystem.

To ensure exceptional quality of product outcomes, our team has adopted a rigorous approach to its innovation process, by using a hybrid combination of the General Statistics Business Process Model (GSBPM) and the International Organization for Standardization (ISO) operating models, as a benchmark for excellence. The IFP's pioneering role in Data Science serves as a valid example for governments and businesses seeking to leverage such platforms, propelling innovative growth by accessing the expansive reservoir of public data and statistics.

Introduction:

Emerging innovations, including Data Science and Big Data are revolutionizing the field of Statistics and Data Science, offering new opportunities to collectively gain insights from Data[1], this will potentially empower industries such as healthcare, autonomous vehicles, smart cities, and supply chain optimization, where the possible applications with Data Science seem almost boundless. Statistical bodies could play an active role in ensuring that these technologies are properly implemented and governed, especially in the Big Data domain. The advent of Big Data has spurred huge advancements in Data Science technologies, allowing us to accelerate in sentient-level decision-making[2], where the combined use of Data Science and Big Data has further potential to revolutionize the way we gather, analyze, and use Data in our organizations.

The growing pool of Data Sets requires innovative methods to effectively analyze and understand the available piles of information. In this paper, we will delve into the benefits of emerging technologies in Statistics and analytics, including Data Science, which has been a key player in this journey. We will also discuss how we use massive amounts of unstructured Data that is available on the internet, and which would have been alternatively time-consuming to analyze by using the traditional methods.

---

[1] (Han, J., Pei, J., & Kamber, M., 2011).
[2] (Cukier, K. N., & Mayer-Schönberger, V., 2013)

Our recent operational adaptability has expanded our role as a National Statistical Office, as We've transitioned from merely publishing descriptive statistics on our website to proactively enabling our stakeholders with data-driven prescriptive insights. This approach answers the four main dimensions of the decision-making practice: 1) What has happened 2) Why did it happen 3) What could happen and lastly 4) How can I make it happen?[3] Altogether, these could be considered a complete toolkit for decision-makers in their day-to-day operations, and further support them in their visionary decision as they better shape the future.

In alignment with the Statistical Revolution and the Strategic Vision of Abu Dhabi, SCAD has developed IFP solely to transform Data into strategic assets, providing users with analytical tools that enable them to diagnose trends and relevant indicators. From this angle, our sole objective is to provide Data-Driven and User-Centric Insights that complement Statistical outputs from SCAD, and that is by enhancing the coverage, timeliness, accuracy, and consistency of our products. In this process, we generally have five sources of Data: 1) Abu Dhabi government entities, 2) the private sector 3) administrative registers, 4) results of survey's & polls and most importantly, 5) Big Data. Our platform is designed in a way that underscores our commitment to aiding the decision-making processes of our primary stakeholders, such as our leadership and government officials.

The state-of-the-art platform (IFP), geared with many analytical tools, delivers high-quality insights analysis such as forecasting Gross Domestic Product (GDP), estimating Small Area Population, nowcasting Real Estate prices and provisionally defining Labor Market characteristics through interactive dashboards. This practice is frequently carried out in collaboration with our Core Statistics team and other partners in the field. Our goal is to continue thriving as Pioneers in the application of Data Science and enablers to our stakeholders, which we aim to maintain by continuously engaging with synergies, amplifying our Data, enriching our Statistics, and generating insights for the betterment of the Emirate.

Insights and Foresights Platform: A One-Stop Shop for Data-Driven Decision-Making in Abu Dhabi

---

[3] (Burstein, F., 2018)

SCAD has pursued and implemented significant initiatives to tap into its potential and support informed decision-making solutions. A prominent stride in this direction is the inception of IFP as a one-stop-shop, serving as a "single source of trust" for Abu Dhabi's strategic agenda, particularly concerning its Economic and Demographic facets. To realize this vision, we have harnessed the power of Data Science and Data-analysis technologies and provided our constant support to policymaking and future planning in the Emirate. From its inception, IFP was architected to be a user-friendly and interactive environment, offering real-time dashboards, analytical models, and scenario simulations to our users. Such features empower decision-makers and planners to understand Data Sets and their dimensions.

Moreover, the platform presents insights in a visually engaging way via a visually immersive native user-interface. With its capability to assimilate and integrate data from diverse sources, the platform also allows for the efficient use, management, and integration of resources across Abu Dhabi's Statistical Ecosystem. This capacity further boosts the worth of our products and ensures their optimal utilization.

Data Sources: Building a Comprehensive Picture

As mentioned above, the IFP is designed to gather Data from five primary sources: Abu Dhabi government entities, the private sector, administrative registers, surveys and polls, and Big Data. These diverse Data Sources are harnessed through advanced algorithms that have automated Data collection functions, guaranteeing that our analytical models adapt dynamically to new Data sources and our platform evolves around sophisticated analytics and advanced reporting tools.

Our ongoing Data Exploration endeavors have been a key factor to the success of our platform, as we persistently uncover new data reservoirs of supporting Data Sets on a daily basis. Our dedicated team remains proactive in engaging in effective dialogues with our stakeholders to refine the models that we have built and improve the platform's functionality from an end-user perspective.

Effective Collaboration with the Private Sector: The Two-Way Road and Our Success Story

SCAD continuously collaborates with the private sector, as this robust partnership consistently yields positive outcomes to our harmonized Statistical Ecosystem. Through sustained engagement, we have pinpointed and liaised with various synergies in the field, notably in 2022, when we entered into several Memorandums of Understanding and Service Level Agreements with Government Entities and private sector firms (acknowledged as pivotal drivers in our domestic economy).

Understanding the significance of these firms and their future outlook has extensively helped us draw a clearer picture with our Insights and Foresights and significantly refined our analysts' perspectives pertaining to market dynamics and trends.

In addition to empowering the government, the platform could also be a valuable asset for the business community in the near future. With access to accurate Statistics and informative Data, they will be able to confidently plan strategies that benefit both the private sector and the economy as a whole in the long run[4], as the platform furnishes extensive Data Sets, visualization tools, and models that enable users to analyze the potential impact of the proposed policies across a spectrum of sectors. In parallel, it could also support comparative benchmark studies of Abu Dhabi with similar economies, augmenting the repository of insights that our users can tap into.

From Exploration to Prediction: Our Data Collection and Analysis Process

In our practice, we make sure to initiate our Data Collection processes by exploring all our options before taking any steps in formulating Data ingestion plans, address both potential challenges and opportunities in data collection. Typically, this follows consultations with pertinent stakeholders to ensure comprehensive coverage relevant to our products. After that, we collect and prepare Data using formats and methodologies compatible with our model's ingestion concept. Once this stage is finalized, Data is subsequently integrated into our database, where it is analyzed by our inhouse Data Science team. Furthermore, we undertake an extensive data analysis process, which involves the exploration and analysis of Data Points and descriptive Statistics. This step also involves reviewing the standard statistical measures, such as

---

(Burstein, F., 2018)
Lee, D., 2021)

outliers, bell curve distribution, mean, and so forth to gain a clearer understanding of the underlying statistical characteristics in Data Sets.

Concurrently with the analysis process, Data Sets are also checked visually using interactive charts, which helps the team in better understanding and identifying trends that may not be immediately clear from granular Data. Through this process, the team can provide a comprehensive view of Data Dimensions with supportive insights. The IFP team uses advanced Data Science methods to formulate insights from Data Sets and create predictive models that are technically robust and realistic, where our models incorporate Data Science via auto machine learning algorithms. The team's profound expertise in data analysis and diverse domains enables them to uncover new insights and identify trends that are hidden within Data.

Understanding Data at a Glance: Benefiting from IFP's Analytical Applications

Before presenting our findings, our products undergo a thorough visualization process, wherein we craft a user interface that effectively presents some of the most essential Data Characteristics. As introduced earlier, IFP's Dashboards and Analytical Applications serve as intuitive graphical tools that are designed to help users analyze and comprehend key indicators. These techniques are specifically crafted to offer a clear and concise view of our Insights and Foresights, enabling users to make informed decisions on the spot. Tools integrated within the IFP are pivotal for our users during their strategic planning, as they continuously seek to gain a deeper understanding of the Data they analyze. Insights generated by our team and presented through IFP have significant importance for our users, and we consistently refine this approach to derive enriched and more informed policies and visionary strategies for Abu Dhabi's Leadership.

A Rigorous Approach to Innovation

We are highly invested in maintaining excellence in our operational framework and have adopted a rigorous approach to our innovation, utilizing a hybrid combination of the GSBPM and the ISO 56002:2019 operating models as our benchmark. This process has been adapted to comply with our specific needs, enabling us to develop use cases that meet the highest standards of quality and precision.

The innovation process consists of five distinct phases, each of which is essential to deliver a successful use case:

I. **Specifying User Needs and Opportunity:** the focus in this first phase is to understand and gather the needs of our stakeholders and users. This includes conducting market research, vendor interviews, and analyzing stakeholders' feedback to pinpoint the areas of enrichment, requirements, and practical preferences. Once the user needs are identified and validated, they are prioritized based on their importance and impact on the business objectives. A complete business case is prepared based on the requirements and strategic objectives, outlining the use case's scope, deliverables, and strategic benefits. The goal of this phase is to ensure that the use cases are aligned with the needs and preferences of stakeholders.

II. **Designing Product Solutions:** this refers to the process of creating a plan for implementing a use case. It may involve proposing a methodology or concept for designing models, reviewing, and refining design concepts, and producing a final document outlining the technical details, workflows, and architecture of the solution. The purpose of the design solution document is to affirm the clarity and definitiveness of the execution plan for the use case.

III. **Developing Product Solutions:** The Data Science team identifies solutions to be used in the Data Preparation, Data Engineering, and a model ingestion process. During this phase, Data Outputs undergo rigorous review and refinement to ensure that they meet the highest standards of quality and accuracy. The team works on developing models and algorithms that will be used to analyze and process the Data, as our aim in this phase is to develop high-quality and accurate product solutions that meet our expectations and those of the stakeholders.

IV. **Deploying Product Solutions:** Data Outputs are deployed in the staging environment, and visual outputs are validated through quality assurance exercises. In addition, user acceptance tests are conducted based on specific tests to identify any issues that need to be addressed before deploying a product in the production environment. This helps ensure that the use cases are ready and will deliver the needed outcomes when deployed in the production environment.

V. **Product Maintenance and Enhancements:** In the final phase of our practice, we shift our concentration to support the evaluation process and

product enhancement plans that are prepared to ensure that the use cases continue to deliver exceptional results over time. This may involve monitoring the performance of products, identifying areas for improvement, and updating products to address any issues that are flagged. If enhancements and updates are required for Data and used models, necessary steps are taken to improve and validate the Data Output, and a new model is planned to be delivered to the production environment. The objective is to ensure that the use cases continue to meet the requirements of our stakeholders and deliver the intended outcomes consistently.

In summary, the hybrid model is a pivotal part of our innovation engine at SCAD, as it provides a systematic and effective way to develop, test, and refine our use cases, ensuring that they meet the highest standards of quality and usability[5]. Our commitment to this process helps us consistently deliver exceptional results and maintain the highest standards of quality and accuracy to our users.

Sample of IFP Products:

I.   Small Area Estimation Model (SAE)

The population census exercise is a traditional Statistical practice that aims to obtain near-realistic population demographics; however, the downside of frequent census exercises is their costliness and time-consuming nature. National Statistical Offices around the world often face challenges in estimating population (as mandated by them), due to irregular and limited availability of important administrative Data such as identity Data. To overcome this challenge, SCAD has modernized the Statistical production models by utilizing alternative Data Sources and Big Data capabilities to develop an in-house Data Science model that makes this exercise effective and less resources intensive. Our modern and innovative approach to capturing the population estimate is called the "Small Area Estimation (SAE) model," which estimates the total population of the Emirate and precisely delineates its distribution across regions and districts.

---

[5] (International Organization for Standardization, 2019)

Accurate population estimates are crucial to sustain informed policymaking and future planning. Our aim in building this model is to provide an accurate and timely population estimate, by using utility consumption and mobility Data to map out the population distribution by district in Abu Dhabi.

The SAE model analyzes rapid changes in population allocation, resulting from migration and urbanization. The population of non-citizens is generally estimated based on the total population and the population estimates of citizens, this exercise could be considered difficult to capture using the traditional census exercise (every 10 years). In addition to utilizing masked Data on people's location (obtained through geospatial mobile applications), our SAE model refined population estimates down to areas as specific as approximately 100 square meters and in some cases by residential unit, providing an unprecedented granularity of information and quality. The Small Area Estimation (SAE) model employs both machine learning and hybrid models' utilization of consumption data and supplementary sources. The collected data is mainly stemmed from the population data base, SCAD's Labor Force Survey and Household Income & Expenditure Survey. Admin data in the form of water consumption data, identity data, health insurance, and school registrations are utilized to adjust demographics. Data from the education sector, Human Resources entities, Emiratization, and labor worker accommodations also feed the SAE model. The SAE model operates through the estimation of the Emirate's total population using water consumption data and then breaking it down into smaller units by district, adjusting and enriching the data, based on the available administrative data. To calculate the total population, the model uses the monthly water consumption on a premise level, tailored specifically for residential properties, as they are the primary focus. A 12-month moving average is applied to the water consumption to account for seasonality, while average consumption per person is calculated as the average consumption per household divided by the average household size. The model merely then requires the application of the correctional factors to adjust for biases from the input data.

II.   Basket of Goods Dashboard

Monitoring consumer price changes of essential food and household items is essential to understand their impact on the day-to-day lifestyle of individuals. Tracking the prices of items that are most relevant to households is a particularly useful tool to support decision-makers in assessing inflationary characteristics on consumption-demand and provides decision-makers with accurate and up-to-date information, to assist them in making data-backed decisions and better target financial support to families based on households' consumption.

Amid the unforeseen global pandemic, a major challenge rose due to the unavailability of Consumer Spending Data from official sources. This has created a gap in the available information on households' consumption and an enhancement opportunity to better assess the effect of the pandemic on society as a whole. In order to fill in this gap proactively and address the needs of our users, the team has leveraged our innovative capacity to create a dashboard that pulls information on household consumption behaviour from the supply side and consolidated it all in one dashboard, which was called the "Basket of Goods application".

The Basket of Goods was originally designed to focus on the most essential items purchased by households, filtered by citizenship classification, providing a realistic analysis of goods consumption beyond the traditional Consumer Price Index. In addition, users would be able to analyze household spending habits, serving as a valuable tool for businesses and investors to refine their market strategies and gain insights crucial to their operations.

During this endeavour, we recognized the efficacy of sourcing data from unconventional sources, as it could be utilized to track prices more precisely over a longer period of time. To gather Data using this approach, we identified several methods of preference, such as 1) Data procured directly from merchandise sellers, or from a proxy, giving us a wider Data Range 2) Other regulating and official bodies that have more granular information 3) Web scraping from publicly available websites.

Conclusion:

The combination of Data Science and Big Data had a clear role in accelerating the way we gather, analyze, and utilize our Data. As part of the Emirate of Abu Dhabi, SCAD has recognized the importance of Data as a strategic asset and proactively developed the IFP to harness this vast potential and support informed decision-making solutions. Complementing this vision, our function has adopted an agile approach to ensure the quality of innovation by employing the GSBPM and ISO 56002:2019 as our hybrid business model and the gold standard for excellence.

Through the IFP we have continuously explored and integrated Data from various sources, enabling the efficient use, management, and integration of resources across Abu Dhabi's Statistical Ecosystem. Our platform offers real-time dashboards, analytical models, and simulations that provide a comprehensive understanding from credible information sources specifically address the four pillars of decision-making: what has happened, why did it happen, what could happen, and how can I make it happen.

The platform could stand as a vital resource for the business community by providing an extensive aid of Data Sets, visualizations tools, and models that empower users to delve deep into the potential impact of proposed policies across the spectrum and the areas of their interests. Ultimately, the goal of IFP is to lead in the application of Data Science and generate data-driven and user-centric insights for the betterment of the Emirate. By achieving this goal, SCAD will continue enabling the government with informed decisions and strategically aligned outcomes. This is all summed up in our strategic vision, collectively we embark on the journey to be "driven by Data for a better tomorrow."

References:

Burstein, F., 2018, From Big Data to Big Decisions: A Decision-Making Framework for Policymakers, *Journal of Big Data*, 5(1), 1-14, doi: 10.1186/s40537-018-0151-6

Cukier, K. N., & Mayer-Schönberger, V., 2013, The Rise of Big Data: *How It's Changing the Way We Think About the World, Foreign Affairs*, 92(3), 28-40

Han, J., Pei, J., & Kamber, M., 2011, Data Mining: *Concepts and Techniques, Morgan Kaufmann*

International Organization for Standardization, 2019, *Innovation management system – a practical guide, ISO 56002:2019*

Lee, S., Kim, M., & Lee, D., 2021, Development of an Open Government Data Platform for Data-Driven Governance: *A Case Study of Seoul Metropolitan Government Sustainability*, 13(11), 6011, doi: 10.3390/su13116011