

# Transforming the FBR into a Data Driven Organization – Business Intelligence and Data Mining

*Muhammad Ali Asad Khan\**

\*Pakistan Customs Service, 30<sup>th</sup> MCMC (aliasadkhan@gmail.com)

## Introduction

The role of IT systems is well documented in literature especially the efforts and initiatives taken by tax administrations and the World Customs Organization, such as steps like developing standards for data interchange, cross-border exchange of information, and so on. With advances in information and communications technology and particularly with the new found ubiquity of business intelligence and data mining methods; tax administrations now have a new milestone avenue to achieve and augment the dual objectives of revenue collection and taxpayer compliance through such methods. This paper attempts to examine this idea to explore if Pakistan's Federal Board of Revenue could be transformed into a data driven organization through the use of such methods.

In Pakistan, revenue collection institutions, such as the Federal Board of Revenue (FBR), face complex and tough challenges. Limited budgets, resource and capacity constraints, and other organizational challenges arising from working in a constantly evolving environment characterized by the high expectations of government, private businesses and citizens results in need of continuous new changes and reforms at level of revenue agencies. This change requires adopting new approaches to achieve high performance by delivering the optimum balance of better outcomes with greater cost-effectiveness.<sup>1</sup>

Globally, revenue agencies have moved away from internal, organization-driven business models and processes to risk-driven and, ultimately, relationship-driven approaches. They are also collaborating more with other departments to provide a whole-of-government approach for taxpayer interactions which better serve citizens. Such freedom to innovate comes from constantly evolving technologies. The

---

<sup>1</sup> Accenture. 2007. "Leading Revenue Agencies Employ Risk Management Strategies and Customer-Centric Philosophy to Achieve High Performance, Accenture Report Finds." February 26, 2007. <https://newsroom.accenture.com/industries/health-public-service/leading-revenue-agencies-employ-risk-management-strategies-and-customer-centric-philosophy-to-achieve-high-performance-accenture-report-finds-1.htm>.

technology underpinning these new models and processes is now robust enough to support a holistic and insightful view of each taxpayer, innovative outreach and education programs, and seamless integration of service delivery - all of which will be the hallmarks of public service value and high performance in the future.<sup>2</sup>

At the heart of all such efforts is data, which the FBR has in abundance. Many tax authorities are developing sophisticated data collection platforms to match and exchange taxpayer information in view of the rising demands for tax transparency by governments and supranational organizations (having power or influence that transcends national boundaries or governments). Data mining is then used to mine this data to help increase the collection of taxes, target enforcement measures and boost performance overall (Pasha 1994).<sup>3</sup>

Currently, the process of converting data into information at the FBR is arduous and error-prone. Customs officers are unable to “check” a taxpayer’s activities on the domestic taxes side, which is administered by the Inland Revenue Service and vice versa. This situation creates inefficiencies resulting in lost opportunities of enhancing revenues and the tax base, resultantly increasing the tax-GDP ratio and conducting more robust risk-based audits. It is nothing short of a national economic security imperative to rectify this situation – the FBR needs to change.

## **Statement of the Problem**

The problem identified above could be summarized as that the FBR has currently no cohesive, integrated IT systems infrastructure to capture all its data, presents it for analysis and has no data mining/business intelligence/analytics capabilities. The FBR is not able to emerge as a high performing revenue organization in the absence of these capabilities. It must change.

This paper examines whether Pakistan’s FBR, which is plagued by an archaic IT infrastructure, could make the quantum leap to high performance by employing data mining tools and techniques.

---

<sup>2</sup> Ibid.

<sup>3</sup> Pasha, H. A., & Iqbal, M. A. (1994). Taxation reforms in Pakistan. *Pakistan Journal of Applied Economics*, 10, 47-75.

## **Significance and Scope of the Study**

Tax collection continues to be a challenge in Pakistan. Our tax structure is characterized by a small tax base, significant evasion and avoidance, an inordinate amount of exemptions and concessions (tax expenditure for FY2020 has been estimated at Rs. 1,149 billion), regressive tax policies, and reliance on indirect taxes. The situation is compounded by resource and capacity constraints at the FBR, which struggles to meet revenue targets. Resultantly, Pakistan has suffered from a chronically low tax-GDP ratio over the years. In FY2006 taxes were 8.7% of GDP. In FY2019, taxes crawled up to 10.1% of GDP.<sup>4</sup> Considering that the government relies on these critical funds to service its citizens, it is imperative to use every avenue and technological resource available to enhance revenue collection. The scope of this paper is limited to looking at FBR's IT systems and how BI & DM capabilities can augment revenues. It does not include:

- A discussion on the politics of decision-making, risk analysis, stake-holder engagement, etc. that such a change requires.
- A detailed project/resource plan on business process re-engineering, change management, and implementation.
- Mathematical or technical explanations of the data mining methods suggested in the paper – it is not a primer on statistics, modeling, simulation, linear/non-linear optimization, etc. However, certain concepts have been explained in footnotes, where required.

## **Research Questions**

In light of the foregoing, the fundamental problem statements that this paper will address are:

1. Can big data be used to transform the FBR into a data-driven organization?
2. How can data analytics and data mining methods be used in the FBR, given the nature/characteristics of the vast store of transactional data it has?

---

<sup>4</sup> Economic Survey of Pakistan, 2019-20. Ministry of Finance. Government of Pakistan.

## **Methodology**

This research has used the qualitative method as it is descriptive and analytical in nature, relying on a qualitative analysis of facts, arguments, literature, policy, and strategy documents, BI tools, and DM methods. Quantitative methods are not used as this paper does not cover statistical modeling techniques used in DM.

## **Organization Of the Paper**

The paper is organized as follows:

- **Section 1**, introduces concepts of data mining and how organizations, including those in government, benefit.
- **Section 2**, provides an overview of the current state of FBR, involving a historical overview of reforms at Pakistan, early reforms and changes needed for FBR.
- **Section 3**, makes a case for change and presents the desired state that the FBR should be in by adopting the use of data mining methods and their benefits. It involves looking at case studies and use cases that have demonstrable impact on revenue and performance. It will discuss various data mining use cases, techniques, and methods that can be used to substantial benefit.
- Conclusion & Recommendations.

## **Review of the Literature**

There is tremendous literature on how to use data mining in a business context. For instance, Linoff and Berry's textbook gives detailed examples on how for-profit businesses use data mining tools and methods in marketing, sales, and customer relationship management to increase revenues, reduce margins, improve customer life cycle management through various techniques, such as hypothesis testing, directed, and undirected data mining algorithms (Linoff 2011).<sup>5</sup> However, for the purposes of this paper, it would be more pertinent to look at literature and examples of how revenue agencies have used these methods.

---

<sup>5</sup> Linoff, Gordon S., and Michael J. A. Berry. 2011. *Data Mining Techniques: For Marketing, Sales, and Customer Relationship Management*. 3rd ed. Indianapolis, IN: Wiley Pub.

The International Monetary Fund (IMF) recognizes that investment in IT and knowledge management, incorporating third-party data is critical for revenue mobilization and improving tax compliance (International Monetary Fund 2015).<sup>6</sup> The IMF's report also identifies that leveraging big data is a key challenge for revenue administrations and successful revenue administration depends upon managing it effectively. It further identifies opportunities where well-designed IT systems can increase the efficiency of tax officials by automating various tasks, such as pre-population of tax returns, verification checks and cross-referencing with third party databases and augmenting these tasks with data mining (International Monetary Fund 2015).<sup>7</sup>

However, there are numerous examples of countries such as Kenya, Nigeria, and even of revenue administrations in developed countries, such as the IRS in the United States that faced difficult challenges in terms of adopting, adapting, capacity building, and scaling such new technologies. Therefore, there is a need for a comprehensive gap analysis, planning, implementation strategy, change management, and monitoring and evaluation mechanisms to properly implement the adoption of such technologies in a holistic design that interweaves technology and tax policy and systems, noting the benefits of such an approach, including reduced interaction with the public and cutting opportunities for rent-seeking behavior on the part of tax officials (Bird 2020).<sup>8</sup>

A summary report on the Lisbon Tax Summit, 2018 held under the auspices of the Inter American Center of Tax Administrations and the Intra-European Organization of Tax Administrations at which the revenue administrations of 80 countries gathered sheds tremendous insights into how revenue administrations around the world are using data mining. For instance, it notes that the Office of Data Science (DS) was established in the United States in 2017 that lead to more robust enforcement and compliance initiatives.

It also notes the Indian experience, where through the adoption of "blockchain technology with big data and advanced analytics offer enormous opportunities for tax administrations to enhance compliance enforcement operations at the national level (tax payment processing, land and population cadasters, etc.) as well as at the

---

<sup>6</sup> International Monetary Fund. 2015. "Current Challenges in Revenue Mobilization - Improving Tax Compliance." Policy Papers 2015, no. 5. <https://doi.org/10.5089/9781498344890.007>.

<sup>7</sup> International Monetary Fund. 2015. "Current Challenges in Revenue Mobilization - Improving Tax Compliance." Policy Papers 2015, no. 5. <https://doi.org/10.5089/9781498344890.007>.

<sup>8</sup> Bird, Richard M., and Eric M. Zolt. "Technology and Taxation in Developing Countries: From Hand to Mouse." *National Tax Journal* 61, no. 4 (2008): 791-821. Accessed November 27, 2020. <http://www.jstor.org/stable/41790481>.

international level (AEOI, country-by country reporting, beneficial ownership, etc.)." Such initiatives enabled India to raise INR 10 billion in additional revenue over 2016-18. It also notes the Belgian example involving the "usage of social networks to predict fraudulent insolvencies, machine learning to interpret data, chat boxes to improve assessments and an on-line portal ("My Minfin"), whereby tax returns can be filed, assessments can be notified, and information on real-estate can be found (Tax Administration & Challenges of the Digital World 2018)."<sup>9</sup>

Therefore, the literature establishes the tremendous gains that can be harvested from the use of big data and data mining methods to increase revenues, enhance compliance, and reduce the costs of compliance for tax-payers. It also points out that such initiatives need to be undertaken with thorough and comprehensive planning to reap the full benefits of such a transformation.

## **Section I**

### **1. Introduction to Data Mining and How Organizations Benefit**

This section establishes a working definition of business intelligence and data mining. It explains their significance, need, and how organizations benefit from it. It also explains the data mining process. While the definitions discussed below generally refer to private organizations, the basics also apply to public sector entities that can leverage their data to improve service delivery and outcomes. For the FBR, such outcomes mean increased revenues, a larger tax base, a better tax to GDP ratio, and so on.

#### **1.1 Concept of Business Intelligence**

Business Intelligence (BI) is defined "as the ability of an organization to take all its capabilities and convert them into knowledge. This produces large amounts of information which can lead to the development of new opportunities for the organization".<sup>10</sup> This necessarily involves using the use and analysis of large amounts of data. When such opportunities have been identified and a strategy has been effectively implemented, they can provide organizations with competitive advantage and stability in the long run within the industry.<sup>11</sup>

---

<sup>9</sup> "Tax Administrations and the Challenges of the Digital World." Summary Report. Lisbon Tax Summit, October 24-26, 2018.

<sup>10</sup> *Business Intelligence & Data Mining With SAS Suite. What Is BI?*, presentation/instruction by Beibei Li, Professor. Heinz College, Carnegie Mellon University, Pittsburgh. 2014.

<sup>11</sup> Ibid.

## 1.2 Concept of Data Mining

Data mining (DM) has been formally defined as, “a business process for exploring large amounts of data to discover meaningful patterns and rules.”<sup>12</sup> For purely pedagogical reasons BI was defined separately earlier, giving the impression that it is somehow separate from DM. However, most authors consider DM as the term crowning a long history of data analysis, which would include BI as well, starting with when humans first analyzed data long before even the first computer was invented. As such, DM has been known by many names, such as knowledge discovery, business intelligence, predictive modeling, predictive analytics, and so on.<sup>13</sup>

The definition of DM, has various components, which have been explained below:

- ***DM Is a Process:*** DM, being a process is a continuous and iterative effort. DM starts with data, which is then analyzed, leading to a strategy or a set of actions, which consequently create more data that feeds into the DM process all over again. In practical terms, this means that organizations that want to gain competitive advantage and improve their performance view DM as an integral part of their enterprise strategy to understand their clients, processes, and so on.<sup>14</sup>
- ***DM Works Better with Large Data Sets:*** The world has come a long way from the days when Excel was used to extract meaningful information from data that used to be scarce and relatively small (by modern standards). Today, with increased computing power, low storage costs, and increasingly sophisticated data mining algorithms and techniques, it has become easier to analyze large sets of data – in fact it is more useful to do so. Data mining techniques work better on larger data sets compared to smaller ones.<sup>15</sup>
- ***DM Provides Meaningful Patterns & Rules:*** The authors describe this as the most important part of the definition, i.e., the discovery of patterns and rules from data that are useful for the organization. There has to be value at the end of the process.<sup>16</sup>

---

<sup>12</sup> Linoff, Gordon S., and Michael J. A. Berry. 2011. *Data Mining Techniques: For Marketing, Sales, and Customer Relationship Management*. 3rd ed. Indianapolis, IN: Wiley Pub.

<sup>13</sup> Ibid.

<sup>14</sup> Linoff, Gordon S., and Michael J. A. Berry. 2011. *Data Mining Techniques: For Marketing, Sales, and Customer Relationship Management*. 3rd ed. Indianapolis, IN: Wiley Pub.

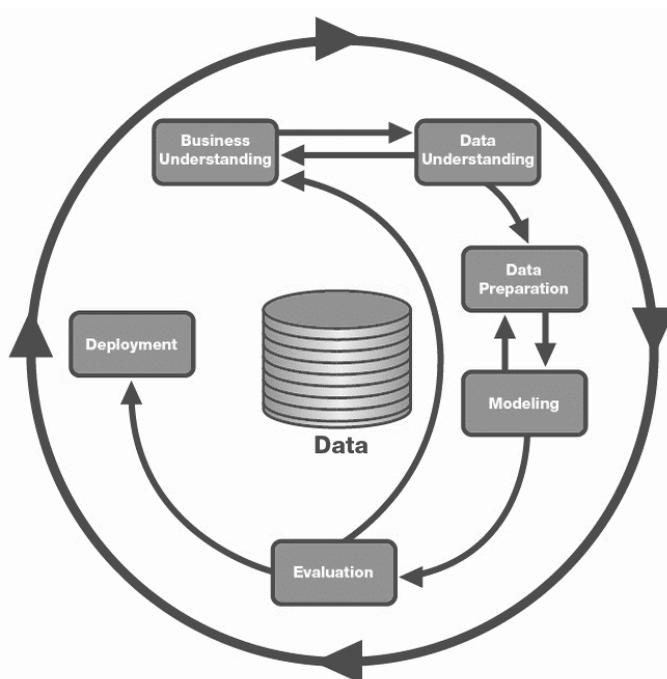
<sup>15</sup> Ibid.

<sup>16</sup> Ibid.

In summary, whether it is referred to as business intelligence or data mining, the discovery of knowledge from vast stores of transactional data that organizations have at their disposal is a process that is used to unearth patterns and relationships that lead organizations to create new strategies to achieve positive outcomes. In this context, at a fundamental level, public sector organizations are no different from the private sector, in that the process of data mining can also be used by the former to achieve organizational objectives as discussed earlier.

All of this can be accomplished under a standard process model that describes common approaches used by data mining experts, known as the Cross-Industry Standard Process for Data Mining (CRISP-DM). The various phases of the data mining process are shown in the figure below and are self-explanatory. Each phase involves a whole host of activities, the discussion of which is beyond the scope of this paper. The key take away here is that the data mining process has a standard set of methodologies and tools that can be used.<sup>17</sup>

**Figure 1: Data Mining Process**



<sup>17</sup> “CRISP-DM by Smart Vision Europe » Data Mining Phases.” n.d. Accessed November 27, 2020. <http://crisp-dm.eu/reference-model/>.

### **1.3 Business Intelligence and Data Mining in the Federal Board of Revenue**

A report by the technology firm Accenture on high performing revenue organizations around the world found that these agencies, relying on IT being a force multiplier, are:

- Using marketing campaigns and other activities to improve public understanding of tax requirements and perceptions of fairness to encourage voluntary compliance. Through this outreach they are creating their revenue "brand."
- Taking an increasingly targeted and proactive approach to collections and compliance. These agencies are using risk models and management strategies to help detect non-compliance and mitigate undesirable outcomes.
- Putting a customer-centric philosophy into practice by providing greater convenience to taxpayers, streamlining the filing and payment processes, and integrating service delivery around taxpayer needs.
- Harnessing advanced customer relationship management processes to dramatically improve both the quality and responsiveness of their service.
- Using rigorous performance metrics and tapping into shared services and outsourcing to continuously refine their operations and redeploy resources for a greater return.
- Improving both processes and technology to support a new services-oriented attitude.<sup>18</sup>

As a result, revenue agencies around the world have seen quantum jumps in their efficiency as measured by various metrics – all as a result of implementing a fully integrated, enterprise-wide solution.

---

<sup>18</sup> Accenture. 2007. "Leading Revenue Agencies Employ Risk Management Strategies and Customer-Centric Philosophy to Achieve High Performance, Accenture Report Finds." February 26, 2007. <https://newsroom.accenture.com/industries/health-public-service/leading-revenue-agencies-employ-risk-management-strategies-and-customer-centric-philosophy-to-achieve-high-performance-accenture-report-finds-1.htm>.

## **Section II**

### **2. The Current State of the FBR**

Now, that we have a basic understanding of what BI and DM mean and how they create and add value, we move on to examine the current state of the FBR's IT systems, with a focus on its business intelligence and data mining capabilities. Before that, it is pertinent to consider the historical overview of the IT reform process at the FBR and its rationale. By necessity, the section examines reforms initiated by Pakistan Customs (dealing with Customs processes and taxes) and those by the Inland Revenue Service (dealing with income tax, sales tax, and federal excise) separately – since these efforts were independent of each other and there was no overall cohesive enterprise-level strategy.

#### **2.1 Historical Overview of Reforms at Pakistan Customs**

Historically speaking, the clearance of cargo in Pakistan used to be a slow, painstaking, and arcane process. Clearing an import cargo required 62–113 steps and 26–34 customs officials. Clearing an export cargo required 31–46 steps and 17–20 handling officials.<sup>19</sup> The average time needed for clearance of goods was more than 7 days, according to various surveys.<sup>20</sup> Customs regulations were complicated and clearance procedures were unwieldy, which gave rise to endless litigation. Some 80,000 cases were pending in courts and Pakistan was one of the leading sources of classification disputes referred to the World Customs Organization for arbitration.<sup>21</sup> Clearly, this left Pakistani trade at a competitive disadvantage.

##### **2.1.1 Early Reforms and Their Failure**

The government attempted to solve some of these problems by introducing reforms of one kind or the other at various times. As part of the reforms, the computerization of

---

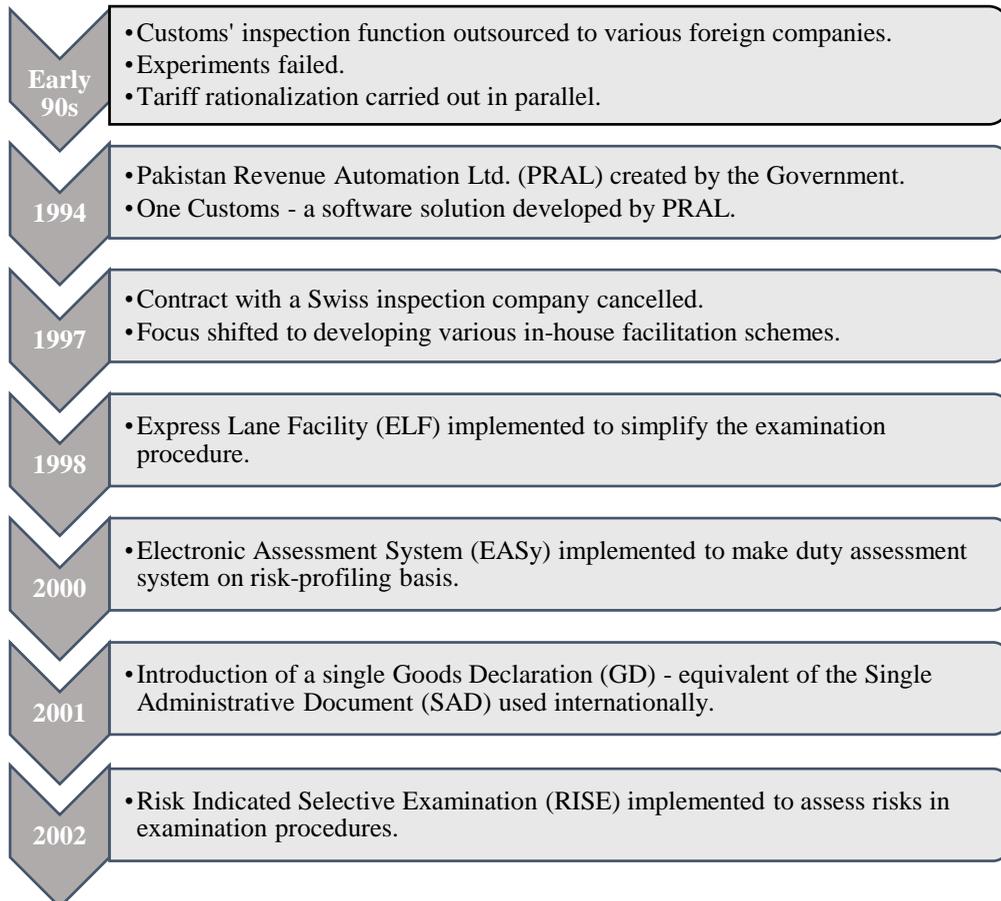
<sup>19</sup> Dennis, Allen. *Trading Across Borders. Case Study: Pakistan. Speeding Up Trade*. Doing Business. The World Bank. <https://www.doingbusiness.org/content/dam/doingBusiness/media/Reforms/Case-Studies/2007/DB07-CS-TOB-Pakistan.pdf>.

<sup>20</sup> *PaCCS – The Electronic Gateway to Customs*, presentation/instruction by Azhar Majeed Khalid, Director General. Directorate General of Training & Research, Karachi. 2009.

<sup>21</sup> *Reforming Trade Facilitation: The Experience Of Pakistan*. Presentation by Ambassador Manzoor Ahmad at the World Bank Trade Logistics Conference on May 5, 2008.

Customs services started in 1979.<sup>22</sup> These early initiatives are captured in the diagram below.<sup>23</sup>

**Figure 2: Timeline Showing Early IT Reform Initiatives undertaken by Pakistan Customs**



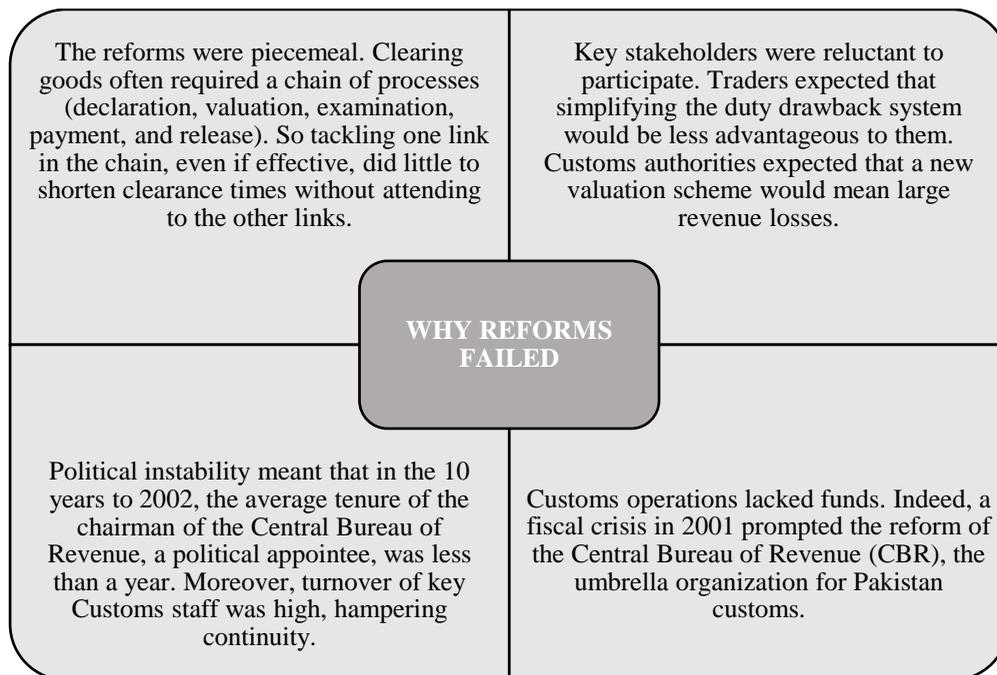
However, despite the best of intentions, these reforms did little to simplify clearance procedures, which continued to be painstakingly slow and cumbersome. One reason was that these reformed procedures were applied selectively only for reputable business establishments, while for the rest clearance time remained unduly long.<sup>24</sup> Other reasons include:

<sup>22</sup> Dennis, Allen. *Trading Across Borders. Case Study: Pakistan. Speeding Up Trade*. Doing Business. The World Bank. <https://www.doingbusiness.org/content/dam/doingBusiness/media/Reforms/Case-Studies/2007/DB07-CS-TOB-Pakistan.pdf>.

<sup>23</sup> PaCCS. *Issues, Challenges & the Way Forward*. Research paper by Muhammad Ali Asad Khan submitted to the faculty of the Directorate General of Training & Research, Karachi in fulfillment of the requirements of the Specialized Training Program. March 2009.

<sup>24</sup> *Ibid.*

**Figure 3: Reasons for Failure of Reforms**



Source: See the footnote<sup>25</sup>

### 2.1.2 The New Era of Reforms – A Paradigm Shift

In this historical context, faced with a fiscal crisis, reforming the tax administration was considered a necessity. The President of Pakistan gave the mandate to an independent task force in June 2000 to review the tax set up and give recommendations on improving it. The task force was headed by Syed Shahid Hussain, a retired World Bank executive. After consulting all stakeholders, i.e., CBR, trade associations, etc. the task force presented its report in May 2001, which “confirmed the inadequacy of the piecemeal approach to earlier customs reforms. And flagging corruption, it identified serious problems in business processes and organization, human resources, and *information management*.” Further, the report stressed that Customs was not to be considered a mere revenue generating agency, but an integral actor in economic development, improving the services of which would lead to a more conducive trading environment, reduced costs for trade, more jobs, economic growth, and a broader revenue base.<sup>26</sup>

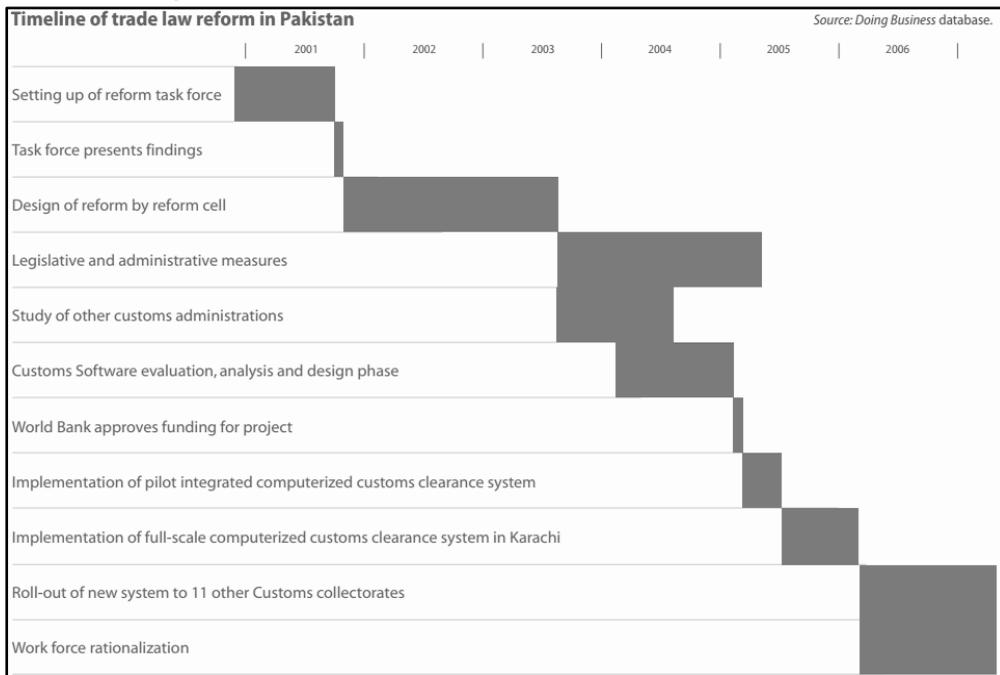
<sup>25</sup> Dennis, Allen. *Trading Across Borders. Case Study: Pakistan. Speeding Up Trade*. Doing Business. The World Bank. <https://www.doingbusiness.org/content/dam/doingBusiness/media/Reforms/Case-Studies/2007/DB07-CS-TOB-Pakistan.pdf>.

<sup>26</sup> Ibid.

Subsequently, in November, 2001, the President of Pakistan approved the Pakistan Tax Administration Reform Program (TARP). The Chairman CBR, Riaz Malik was tasked with creating a reform program. In consultation with the World Bank, International Monetary Fund, UN Conference on Trade and Development, and Maxwell Stamp (a consulting firm), a comprehensive program for CBR reform was approved in March, 2003.

As an integral part of the overall reform program, Pakistan Customs established a Customs Administration Reform (CARE) project, which set out to completely re-engineer customs operations and processes, moving towards 24/7 clearances, self-assessment, risk management, a paperless single window operation, and automation.<sup>27</sup> Under CARE, strategic, comprehensive, and widespread reforms were introduced (see timeline below), a lynchpin of which was to build an online cargo clearance system based on international standards and best practices on Customs, trade facilitation, and integrity.

**Figure 4: Timeline Showing the Next Generation of Reform Initiatives undertaken by Pakistan Customs**



<sup>27</sup> Dennis, Allen. *Trading Across Borders. Case Study: Pakistan. Speeding Up Trade*. Doing Business. The World Bank. <https://www.doingbusiness.org/content/dam/doingBusiness/media/Reforms/Case-Studies/2007/DB07-CS-TOB-Pakistan.pdf>.

Thus, the Pakistan Customs Computerized System (PaCCS) was born, which took about 3 years to be developed. It was the result of an in-house, indigenous research & development effort coupled with Business Process Reengineering (BPR) by Pakistan Customs under the CARE project. Developed by a foreign company, it was an integrated, web-based software solution that provided a common platform to all the major stakeholders – all government and non-governmental entities involved in international trade. Whereas, PaCCS successfully automated and expedited Customs processes, facilitating trade and improving throughput, it had limited MIS functionality and was not amenable to BI and DM tools.<sup>28</sup>

PaCCS eventually outlived its utility and in association with PRAL, Pakistan Customs developed a completely indigenous successor – Web Based One Customs (WEBOC), which was rolled out in 2011 at Port Qasim, Karachi after rigorous testing and eventually throughout the country.<sup>29</sup> Later in 2018, it was upgrade to WEBOC Glo and it is currently operational across the country.

## **2.2 Historical Overview of Reforms at Inland Revenue**

Digitization of business processes at Inland Revenue started in early 2003, when the Medium and Large Tax Payers Units were set up. A number of software systems were developed over the years:

- **Income Tax:** Tax Management System (return entry; refund enforcement), Mahasil, and Iris, which is currently operational
- **Sales Tax:** STARR, CREST, STRIVe, ERS, FASTER, etc.

The problem with this development process was that it was not cohesive or integrated. The various systems developed were disjoint, stand-alone, and monolithic, with extremely limited integration abilities. Further, most could not be updated over time and others abandoned, resulting in a very complex and archaic IT systems infrastructure. Currently, Inland Revenue has the following systems in operation: Iris,

---

<sup>28</sup> PaCCS. Issues, Challenges & the Way Forward. Research paper by Muhammad Ali Asad Khan submitted to the faculty of the Directorate General of Training & Research, Karachi in fulfillment of the requirements of the Specialized Training Program. March 2009.

<sup>29</sup> “‘WeBOC’ Clearance System Launched at PQ.” 2011. DAWN.COM. June 3, 2011. <https://www.dawn.com/2011/06/03/weboc-clearance-system-launched-at-pq/>.

FASTER & ERS, STRIVE, Paysys, CRM, Point of Sales, Tax Asaan App, Operational Data Source with Oracle BI, Dashboards for senior executives, E-payment systems.<sup>30</sup>

### **2.3 Current State of FBR's Data Analytics Capacity**

Since 2014 onwards, FBR has been maintaining Data Warehousing solution in the form of Operation Data Store (ODS), which is a central repository of data directly captured by FBR systems for Inland Revenue and Customs. Analytics is primarily done at two levels:

1. Centralized analytics is done primarily on the data stored in the ODS by a dedicated team of PRAL.
2. Localized analytics is done in Field Formations. Each formation has DBA who fulfill all analytics related needs of the field formations.

However, the analytics provided under this framework are more akin to rudimentary MIS reports, rather than advanced data analytics, as the required skillset is non-existent. The FBR's own ICT plan admits that, "given the current and future requirements of Inland Revenue, the current infrastructure and human resource is inadequate and does not fit into the transformation Plan which IRS is aiming at."<sup>31</sup>

### **2.4 Critical Analysis**

The ICT Transformation plan developed by the IT Wing, FBR<sup>32</sup> recognizes the issues and challenges of the current IT infrastructure. That is, there is no formal system for information processing, reporting, and advanced analytics. This is primarily due to a lack of ownership, funding, and an integrated, enterprise-wide approach to IT systems development. For the future, the plan also recognizes the importance of analytics. However, once again the plan is myopic as the future IT architecture envisioned only involves Inland Revenue components and does not highlight formal linkages with Pakistan Customs' IT systems or third-party systems, such as commercial banks, land record, excise, etc. This lack of a holistic, all-encompassing, enterprise-wide IT strategy will not alleviate structural problems, such as the fact that the tax

---

<sup>30</sup> ICT Transformation Plan. Historical Perspective, Current Status and Way Forward. September 14, 2020. IT Wing, Federal Board of Revenue, Islamabad.

<sup>31</sup> Ibid.

<sup>32</sup> ICT Transformation Plan. Historical Perspective, Current Status and Way Forward. September 14, 2020. IT Wing, Federal Board of Revenue, Islamabad.

administrators do not have a single view of a tax payer across all taxes or other related departments.

## **Section III**

### **3. Transforming the FBR**

This section examines why the FBR must change to become data-driven, what benefits can be gained, and how it can be done. Starting with the why, the FBR's non-existent data mining capabilities, outdated IT systems, lack of a holistic, enterprise-wide data view, and other structural issues have resulted in a number of problems:

- There is no enterprise-wide knowledge base on tax payers, e.g. income tax officials auditing a person's income tax returns have no means to access the person's customs history in a timely manner or cross reference it with third party databases.
- Data is fragmented, contained in monolithic, stand-alone systems that cannot interface with one another. No data warehouses or data mining to facilitate business intelligence and analyses.
- There is no way for these systems to communicate with external organizations like banks from which additional information can be collected on tax payers.
- It takes a long time for legislative changes to be incorporated into these systems.
- Compliance and by extension, detection of non-compliance is a major issue since there is inadequate risk assessment and risk mitigation strategies driving these outdated software modules.
- There are serious questions regarding the scalability of these systems into the future.
- Business processes remain plagued with bottlenecks and grey areas that arise from a fundamental inability to align people, processes and technology, caused by the absence of an over-arching, holistic, organization-wide IT integration strategy that aligns organizational strategic imperatives to the activities that it carries out – operational inefficiencies that translate into *lost revenue* and *increased cost of collection*.

- The FBR is engaged in costly and conflict-ridden, internal departmental restructuring efforts aimed at trying to integrate the various taxes around tax types and functions in order to improve collection and compliance without addressing the underlying fundamental problems relating to the taxpayer knowledge and database that is causing problems in the first place.
- Service delivery suffers and taxpayers' grievances go up.

As a result, Pakistan suffers from one of the lowest tax-GDP ratios in the region – a mere 10-11%. We also have one of the lowest tax bases – 1.3% of the population as against 20% in Malaysia and 10% in Turkey. Non-compliance and evasion result in significant cost and time overheads as a result of the additional burden of investigative and legislative procedures. Revenue targets have to be revised downwards as a result of these structural and systemic flaws in the FBR's IT infrastructure. This is why the FBR needs to change in order to become a high performing revenue organization. By adopting some of the BI & DM use cases below (not an exhaustive use case list), the FBR can accomplish just that.

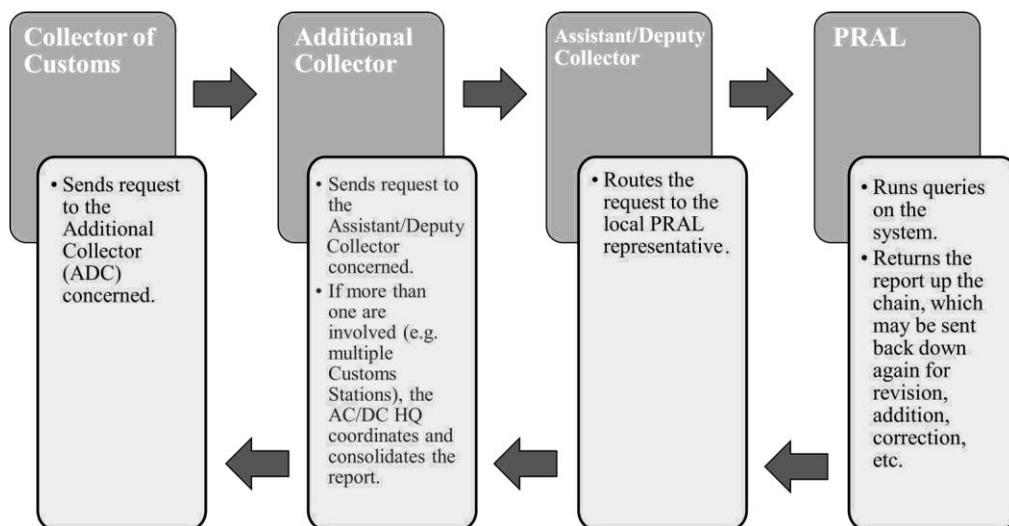
### **3.1 Advanced BI & Analytics Dashboard for Real-Time Information**

Let us consider a typical use case at a Collectorate.<sup>33</sup> The Collector would like a report on the duties collection, broken down by tax head and differentiated by the various Customs Stations under his control. Currently, this request would be routed along the following chain, going through numerous iterations and revisions.

---

<sup>33</sup> A field formation of Pakistan Customs, headed by a Collector of Customs (BS20), primarily concerned with operational and transactional matters, i.e. imports, exports, enforcement, anti-smuggling, etc.

**Figure 5: Chain of Requests for Forming a Report on Duties Collection**



This process is fraught with delays and inefficiencies – it can take up to an entire working day to get the required report, resulting in delayed decision-making.<sup>34</sup> In the context of international trade and supply chains engineered for efficiency, this situation is unacceptable. As a proof-of-concept, the author, while posted in Kohat used PRAL import data to create a real-time BI & Analytics Dashboard that could provide a range of reports, KPIs, and indicators, which would enable the Collector to generate/view such a report on his/her own instantly, save precious time, enable quick decision-making, and directly impact service delivery. Partial snapshots of the data (10,000 records across 48 variables) are given below:

<sup>34</sup> The specific example in this use case reflects any situation where reports based on the data store of the FBR are required. This is just a typical example.

Figure 6

GD Number	GD Date	NTN	Importer Name on GD	Importer Address on GD	Importer Name on IGM	Agent Name
SHAH-HC-3-03-05-2016	03-05-16	1279172	LAHORE STEEL MILL	68 INDUSTRIAL ESTATE JAMR LAHORE STEEL	AL SHUJA & BROTHERS	
SHAH-HC-4-03-05-2016	03-05-16	1358881	TAJ WAZIR ENTERPRISES	BAKA KHEL WAZIR, P/O CUST TAJ WAZIR ENTERPRISES		
SHAH-HC-5-04-05-2016	04-05-16	1358881	TAJ WAZIR ENTERPRISES	BAKA KHEL WAZIR, P/O CUST TAJ WAZIR ENTERPRISES		
SHAH-HC-6-04-05-2016	04-05-16	1358881	TAJ WAZIR ENTERPRISES	BAKA KHEL WAZIR, P/O CUST TAJ WAZIR ENTERPRISES		
SHAH-HC-7-06-05-2016	06-05-16	1358881	TAJ WAZIR ENTERPRISES	BAKA KHEL WAZIR, P/O CUST TAJ WAZIR ENTERPRISES		
SHAH-HC-8-09-05-2016	09-05-16	1358881	TAJ WAZIR ENTERPRISES	BAKA KHEL WAZIR, P/O CUST TAJ WAZIR ENTERPRISES		
SHAH-HC-10-11-05-2016	11-05-16	1279172	LAHORE STEEL MILL	68 INDUSTRIAL ESTATE JAMR LAHORE STEEL MILLS PVT LTD	AL SHUJA & BROTHERS	
SHAH-HC-9-10-05-2016	10-05-16	1358881	TAJ WAZIR ENTERPRISES	BAKA KHEL WAZIR, P/O CUST TAJ WAZIR ENTERPRISES		
SHAH-HC-11-11-05-2016	11-05-16	1279172	LAHORE STEEL MILL	68 INDUSTRIAL ESTATE JAMR LAHORE STEEL MILLS PVT LTD	AL SHUJA & BROTHERS	
SHAH-HC-12-11-05-2016	11-05-16	1279172	LAHORE STEEL MILL	68 INDUSTRIAL ESTATE JAMR LAHORE STEEL MILLS PVT LTD	AL SHUJA & BROTHERS	
SHAH-HC-13-11-05-2016	11-05-16	1358881	TAJ WAZIR ENTERPRISES	BAKA KHEL WAZIR, P/O CUST TAJ WAZIR ENTERPRISES		
SHAH-HC-14-11-05-2016	11-05-16	1358881	TAJ WAZIR ENTERPRISES	BAKA KHEL WAZIR, P/O CUST TAJ WAZIR ENTERPRISES		
SHAH-HC-16-12-05-2016	12-05-16	1358881	TAJ WAZIR ENTERPRISES	BAKA KHEL WAZIR, P/O CUST TAJ WAZIR ENTERPRISES		
SHAH-HC-15-12-05-2016	12-05-16	1358881	TAJ WAZIR ENTERPRISES	BAKA KHEL WAZIR, P/O CUST TAJ WAZIR ENTERPRISES		

Figure 7

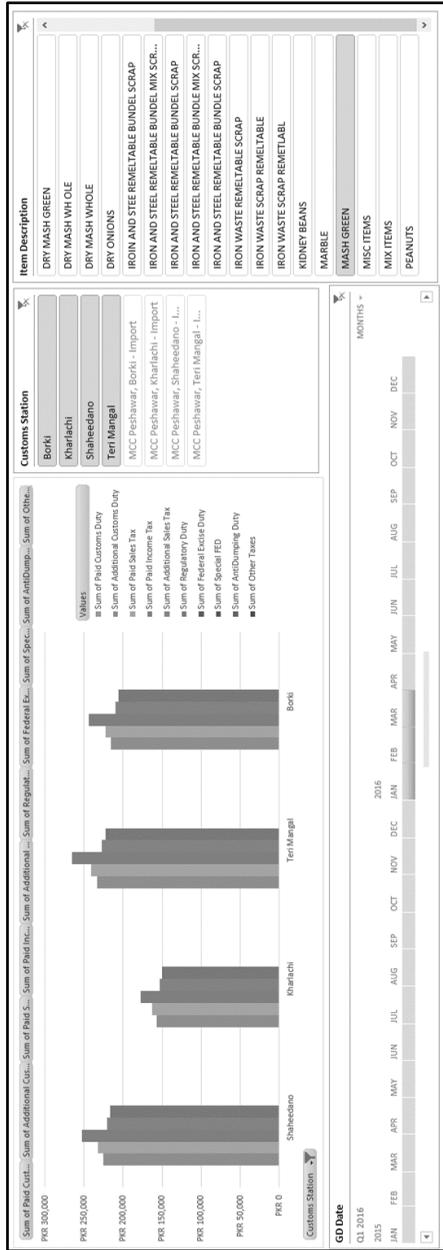
PCT Code	Dec Unit Value	Assessed Unit Value	Currency	Quantity	UOM	Item Description	Import Value in PKR	origin	Released Date	Statutory Customs Duty
7204100	\$	0.21	\$	28,000	KG	IRON AND STEEL REMELTABLE BUNDLE SCRAP	PKR 627,410	Afghanistan	42493	PKR 12,548.00
7133000	\$	0.32	\$	30,000	KG	KIDNEY BEANS	PKR 1,024,344	Afghanistan	42493	PKR 20,487.00
7139020	\$	0.23	\$	30,000	KG	DRY WASH WHOLE	PKR 736,247	Afghanistan	42494	PKR 14,725.00
7139020	\$	0.23	\$	30,000	KG	DRY WASH WHOLE	PKR 736,247	Afghanistan	42494	PKR 14,725.00
7139020	\$	0.23	\$	30,000	KG	DRY WASH WHOLE	PKR 736,247	Afghanistan	42496	PKR 14,725.00
7204490	\$	0.28	\$	28,000	KG	IRON AND STEEL REMELTABLE BUNDEL SCRAP	PKR 836,547	Afghanistan	42501	PKR 16,715.00
7139020	\$	0.23	\$	30,000	KG	IRON AND STEEL REMELTABLE BUNDEL SCRAP	PKR 736,247	Afghanistan	42500	PKR 14,725.00
7204490	\$	0.28	\$	28,000	KG	IRON AND STEEL REMELTABLE BUNDEL SCRAP	PKR 836,547	Afghanistan	42501	PKR 16,715.00
7139020	\$	0.23	\$	30,000	KG	DRY WASH WHOLE	PKR 736,247	Afghanistan	42501	PKR 14,725.00
7139020	\$	0.23	\$	30,000	KG	DRY WASH WHOLE	PKR 736,247	Afghanistan	42502	PKR 14,725.00
7139020	\$	0.23	\$	30,000	KG	DRY WASH WHOLE	PKR 736,247	Afghanistan	42502	PKR 14,725.00
7204490	\$	0.28	\$	28,000	KG	IRON AND STEEL REMELTABLE BUNDEL SCRAP	PKR 835,748	Afghanistan	42507	PKR 16,715.00
7204490	\$	0.28	\$	28,000	KG	IRON AND STEEL REMELTABLE BUNDEL SCRAP	PKR 835,748	Afghanistan	42507	PKR 16,715.00
7204490	\$	0.28	\$	28,000	KG	IRON AND STEEL REMELTABLE BUNDEL SCRAP	PKR 835,748	Afghanistan	42508	PKR 16,715.00

From this data the following *interactive* dashboards were prepared, which extract meaningful information in real-time, allowing non-technical officers to instantly manipulate the underlying data through an intuitive interface consisting of filters, slicers, timelines, and buttons to select various options.<sup>35</sup> For instance, the graph on the

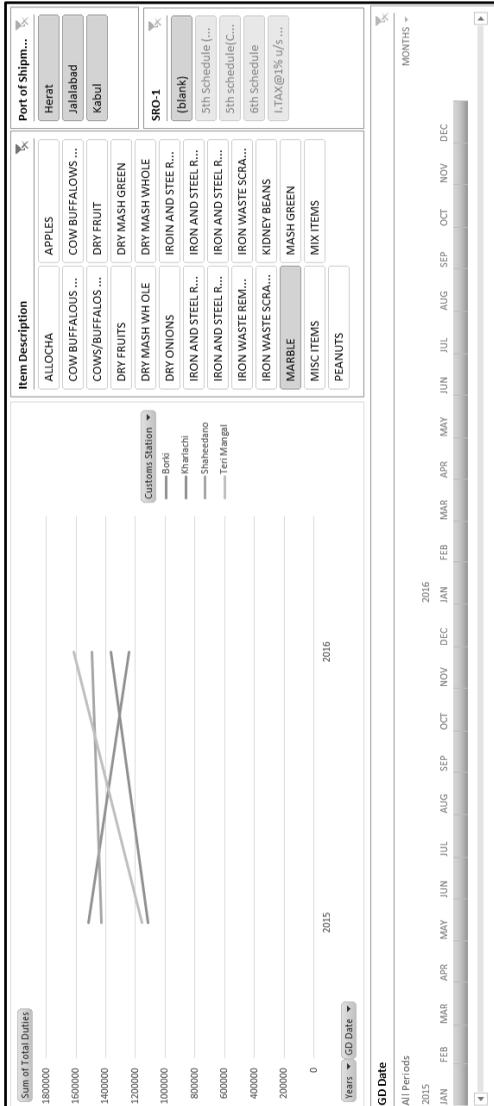
<sup>35</sup> Including the one given in the typical example above.

dashboard below shows duties breakdown by tax on the import of ‘Mash Green’ at four Customs Stations for the period Jan-March 2016 – and it can be instantly modified. The possibilities arising from such efficient manipulation and presentation of information are endless, likely to improve informed decision-making.

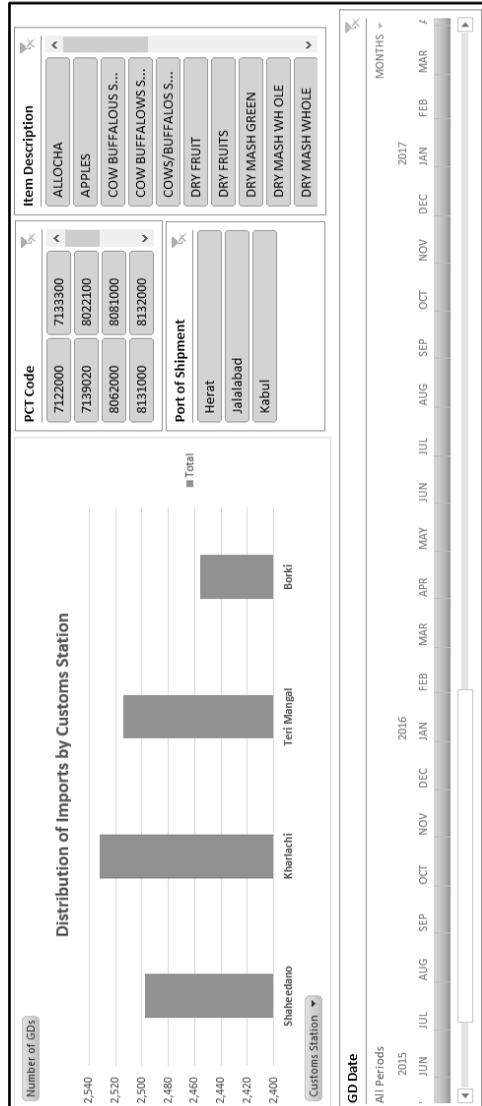
Figure 8



Other BI Dashboards Prepared by the Author are Given Below:



**Figure 9: Dashboard for Analyzing Revenue Trends by Customs Station, Items, Port of Shipment, and Exemptions Granted**



**Figure 10: Dashboard for Analyzing Import Volume by Customs Station, Items, Port of Shipment, and PCT Code**

### **3.2 Using Predictive Analytics to Forecast Revenue**

There are a number of robust, time-tested statistical and time series forecasting methods that the FBR can use to predict future revenues based on historical time series<sup>36</sup> revenue data. These techniques can also be used to predict import/export volumes, demand for imports, income tax revenues, and even the revenues of large businesses – reported deviations from which can drive investigative/enforcement initiatives to counter under-reporting. This is most important for a cash-strapped nation such as Pakistan – it reduces the uncertainty associated with future inflows and enables informed decision-making in critical areas such as fiscal, tax, and monetary policy.

Each of these models listed below range from the simple to the complex, varying in predictive accuracy and in their modeling of stationary, non-stationary, and seasonal patterns<sup>37</sup> in the data. A common approach would be to try multiple methods and evaluate their predictive accuracy based on how well they predict past data, which is evaluated on various accuracy measures, such as Mean Absolute Deviation or Mean Absolute Percent Error. A number of models can also be fed into a composite, weighted statistical model that reduces the error rate and increases predictive accuracy.<sup>38</sup> A brief list of such models and techniques is provided below:<sup>39, 40</sup>

---

<sup>36</sup> A time series is a set of observations on a quantitative variable collected over time.

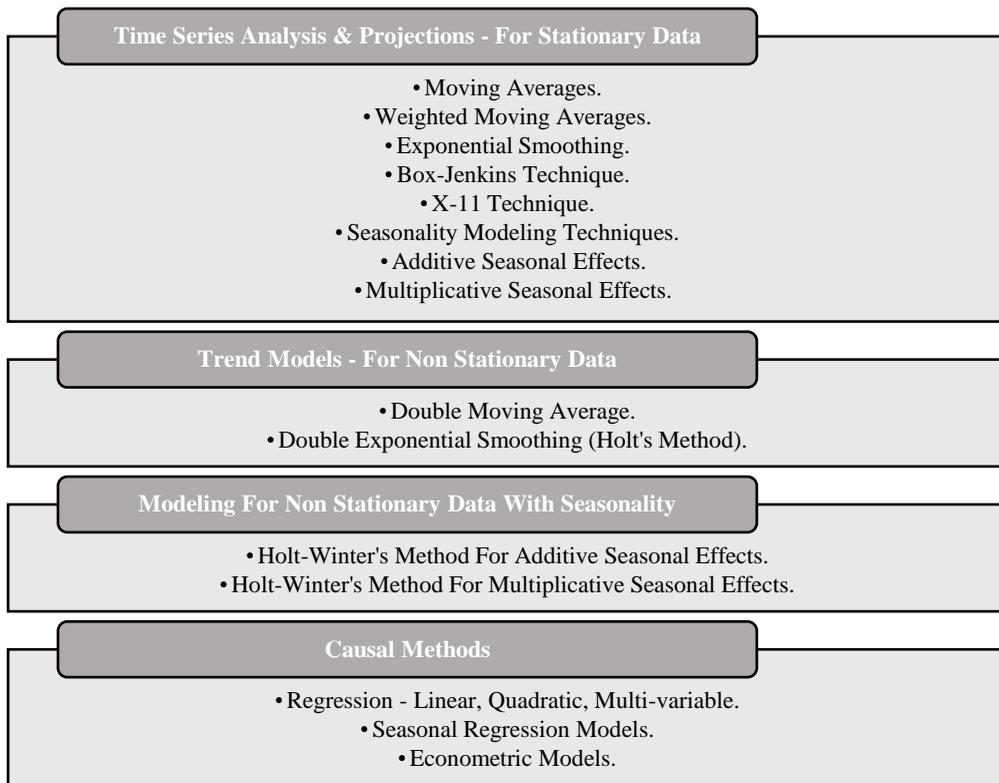
<sup>37</sup> Stationary data-no trends; Non-stationary data-exhibits long term trends; Seasonal-exhibits seasonal patterns.

<sup>38</sup> Ragsdale, Cliff T. *Spreadsheet Modeling & Decision Analysis. A Practical Introduction to Management Science.* Edition: 6. Mason, OH, USA: South-Western, Cengage Learning.

<sup>39</sup> Ibid.

<sup>40</sup> Chambers, John C. Mullick, Satinder K. Smith, Donald D. 1971. *Forecasting: How To Choose The Right Forecasting Technique.* Harvard Business Review.

**Figure 11: Statistical Models and Techniques**



### **3.3 Models to Detect High-Risk Refunds – Case Study of The Australian Tax Office**

By 2012, the Australian Tax Office (ATO) had an advanced analytics capability consisting of predictive risk models, large-scale visualizations, use of third-party data, and social network analysis. Their IT infrastructure could, at that time, handle data for 40 million taxpayer transactions and half a billion data points from third parties annually.

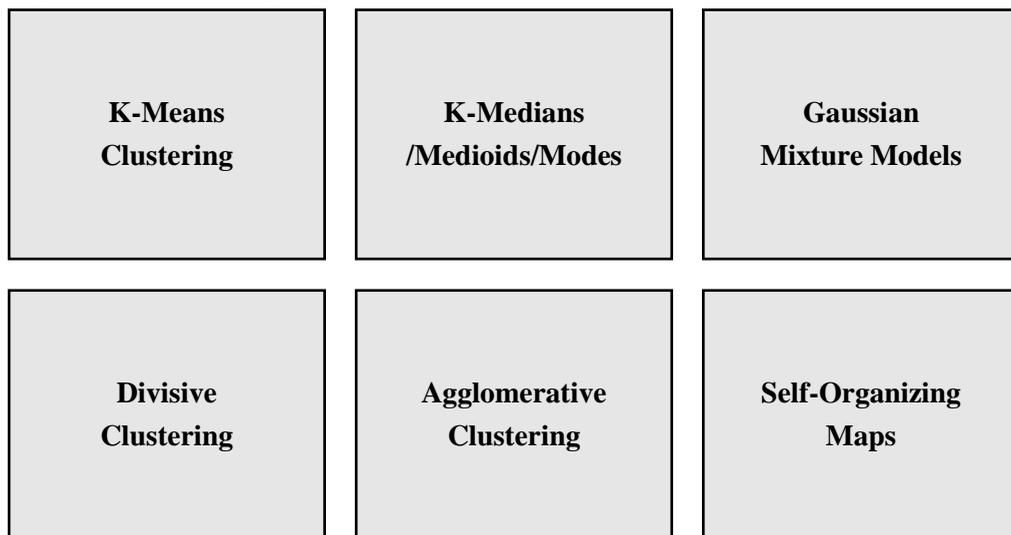
This analytics capability improved the ATO's ability to respond to a taxpayer transaction, based on what is known about the taxpayer holistically. For instance, their models to detect high-risk refunds enabled the ATO to prevent refunds worth \$665 million from being issued incorrectly in 2010-11, based on the models' social network discovery algorithm that detects networks among individuals, companies, partnerships or tax return forms. The models were continuously updated to improve detection of

emerging frauds. This system also helped speed up the processing of legitimate refund cases.<sup>41</sup>

### **3.4 Clustering Algorithms to Better Understand Taxpayers**

There are a number of undirected data mining algorithms<sup>42</sup> that can enable automated segmentation and clustering of taxpayers based on their underlying characteristics (e.g., income, age, education, nature of business, etc.) and transactional behavior (e.g., compliance history, contraventions, exemptions claimed, etc.) that can enable the FBR to better understand the population of taxpayers. This can enable the FBR to group the entire population into relevant clusters and better understand the issues, service needs, and any pre-emptive enforcement actions required for each cluster – an approach that was adopted by the ATO.<sup>43</sup> Some of the clustering algorithms that can be used on the FBR’s data are listed below:<sup>44</sup>

**Figure 12: Clustering Algorithms**



*Source: Linoff, Gordon S., and Michael J. A. Berry. 2011*

<sup>41</sup> Martikainen, Jani. 2012. Data Mining In Tax Administration - Using Analytics To Enhance Tax Compliance. <https://core.ac.uk/download/pdf/80706408.pdf>.

<sup>42</sup> The algorithm creates the clusters and segments “automatically”, possibly revealing important insights, patterns, and groups in the data, which were not considered or thought of a priori.

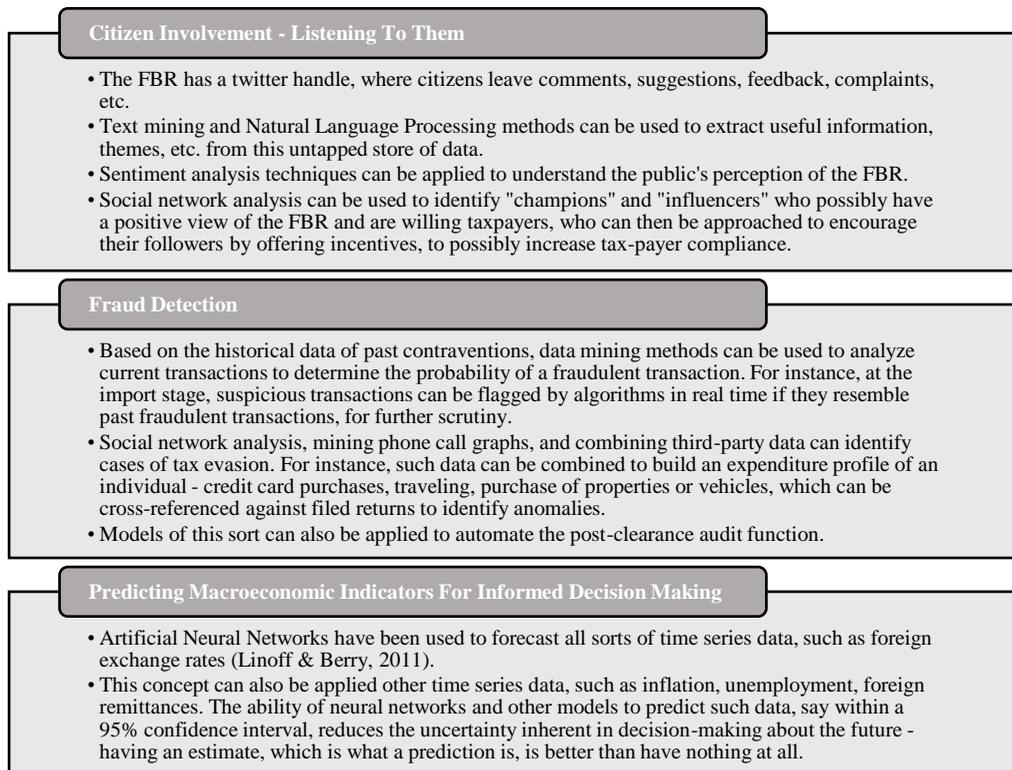
<sup>43</sup> Ibid.

<sup>44</sup> Data Mining Techniques: For Marketing, Sales, and Customer Relationship Management. 3rd ed. Indianapolis, IN: Wiley Pub.

### 3.5 Other Use Cases & Methods

There are many other use cases and scenarios of how the FBR can be transformed into a data-drive organization:

**Figure 13: Cases and Scenarios for Transforming FBR into a Data-Driven Organization**



## 4. Conclusion

Data mining has dramatically reshaped the way governments, businesses, and societies operate and interact with each other. There are tremendous benefits to be achieved by utilizing data mining methods. This paper identified that there is no system for information processing, reporting, advanced analytics, and data mining at the FBR. This is primarily due to a lack of ownership, funding, and an integrated, enterprise-wide approach to development of IT systems. The lack of a holistic, all encompassing, and enterprise-wide IT strategy poses problems, such as the fact that the tax administrators do not have a single view of a tax payer across all taxes or other related

departments. This critically hampers the FBR's ability to collect revenues and this has to change.

This paper has shown that harnessing big data and using data mining techniques can be of tremendous benefit to the FBR, such as: advanced BI and data analytics providing real time actionable insights; forecasting and prediction, for instance predicting revenue trends or trade patterns that can reduce the uncertainty in decision making; enforcement – identifying suspicious transactions; and finally, automated post clearance audit in order to prevent under-collection of revenue.

Going back to the problem statement – in light of the critical discussion presented in this paper, it is reasonable to conclude that the FBR can indeed be transformed into a data-driven, high-performing revenue organization by leveraging data mining tools and techniques, which addresses part one of the problem statement. Section 3 of this paper addresses part two of the problem statement, i.e., it presents use cases, algorithms, and tools showing how data analytics and data mining methods can be used in the FBR and its vast store of valuable data.

## **5. Recommendations**

Based on the analysis and conclusion, the FBR needs to forge a pathway to developing a holistic, integrated IT framework, centered around a data lake that captures in-house and third-party data. This is no small undertaking and requires a well-developed strategic plan, encompassing all elements of project planning, change management, implementation, and monitoring and review.

As a first step, it is imperative that the FBR develop a Data Science wing. The FBR should identify suitable resources within the department and if none are available, they should be hired from the private sector. This Data Science wing should engage in strategic planning in the context of augmenting revenue collection and tax compliance through data mining methods, such as those described in this paper. Such initiatives can be rolled out in pilot programs at selected field formations, their efficacy evaluated based on well-defined KPIs, and if warranted, rolled out to the rest of the country.

It is imperative that the initiative be owned and backed by political and administrative will at the highest levels, develop appropriate and necessary legislation and policy frameworks, such as those related to data privacy and protection, and have a strong monitoring and evaluation component, to ensure a successful implementation.

Since data mining methods have demonstrable impact which can be quantified, the Data Science wing at the FBR and its employees should be adequately incentivized by giving them a percentage of additional revenue generated. This will also help the FBR attract the best and brightest talent. Pakistan needs it.

## **Bibliography**

Accenture. 2007. “Leading Revenue Agencies Employ Risk Management Strategies and Customer-Centric Philosophy to Achieve High Performance, Accenture Report Finds.” February 26, 2007. <https://newsroom.accenture.com/industries/health-public-service/leading-revenue-agencies-employ-risk-management-strategies-and-customer-centric-philosophy-to-achieve-high-performance-accenture-report-finds-1.htm>.

Ahmed, V., & Talpur, M. (2016). Corporate tax reforms in Pakistan.

*Business Intelligence & Data Mining With SAS Suite. What Is BI?*, presentation/instruction by Beibei Li, Professor. Heinz College, Carnegie Mellon University, Pittsburgh. 2014.

Chambers, John C. Mullick, Satinder K. Smith, Donald D. 1971. Forecasting: How to Choose the Right Forecasting Technique. Harvard Business Review.

“CRISP-DM by Smart Vision Europe » Data Mining Phases.” n.d. Accessed November 27, 2020. <http://crisp-dm.eu/reference-model/>.

Dennis, Allen. *Trading Across Borders. Case Study: Pakistan. Speeding Up Trade. Doing Business.* The World Bank.  
<https://www.doingbusiness.org/content/dam/doingBusiness/media/Reforms/Case-Studies/2007/DB07-CS-TOB-Pakistan.pdf>.

Economic Survey of Pakistan, 2019-20. Ministry of Finance. Government of Pakistan.  
Gul, R., & Ahsan, A. (2019, May). Big Data and Analytics: Case Study of Good Governance and Government Power. In *European Conference on Intangibles and Intellectual Capital* (pp. 128-XI). Academic Conferences International Limited.

ICT Transformation Plan. Historical Perspective, Current Status and Way Forward. September 14, 2020. IT Wing, Federal Board of Revenue, Islamabad.

International Monetary Fund. 2015. “*Current Challenges in Revenue Mobilization - Improving Tax Compliance.*” Policy Papers 2015, no. 5.  
<https://doi.org/10.5089/9781498344890.007>.

Lander, Jared P. 2014. *R for Everyone: Advanced Analytics and Graphics.* Addison Wesley Data & Analytics Series. Upper Saddle River, New Jersey: Addison-Wesley.

Linoff, Gordon S., and Michael J. A. Berry. 2011. *Data Mining Techniques: For Marketing, Sales, and Customer Relationship Management*. 3rd ed. Indianapolis, IN: Wiley Pub.

Martikainen, Jani. 2012. Data Mining In Tax Administration - Using Analytics To Enhance Tax Compliance. <https://core.ac.uk/download/pdf/80706408.pdf>.

Martin R, Fardmanesh M. Fiscal variables and growth: A cross-sectional analysis. *Public Choice*. 1990 Mar 1;64(3):239-51.

Myles GD. Taxation and economic growth. *Fiscal studies*. 2000 Mar;21(1):141-68.  
Ndenga, Sidwaka J., and Caro Ayuma. "Effect of Post Clearance Audit on Revenue Collection in Kenya: A Case Study of Customs Service Department in Mombasa County."

*PaCCS – The Electronic Gateway to Customs*, presentation/instruction by Azhar Majeed Khalid, Director General. Directorate General of Training & Research, Karachi. 2009.

PaCCS. Issues, Challenges & the Way Forward. Research paper by Muhammad Ali Asad Khan submitted to the faculty of the Directorate General of Training & Research, Karachi in fulfillment of the requirements of the Specialized Training Program. March 2009.

Pasha, H. A., & Iqbal, M. A. (1994). Taxation reforms in Pakistan. *Pakistan Journal of Applied Economics*, 10, 47-75.

Ragsdale, Cliff T. *Spreadsheet Modeling & Decision Analysis. A Practical Introduction to Management Science*. Edition: 6. Mason, OH, USA: South-Western, Cengage Learning.

*Reforming Trade Facilitation: The Experience Of Pakistan*. Presentation by Ambassador Manzoor Ahmad at the World Bank Trade Logistics Conference on May 5, 2008.

Taha R, Nanthakumar L, Colombage SR. The effect of economic growth on taxation revenue: The case of a newly industrialized country. *International Review of Business Research Papers*. 2011 Jan;7(1):319-29.

“Tax Administrations and the Challenges of the Digital World.” Summary Report. Lisbon Tax Summit, October 24-26, 2018.

Bird, Richard M., and Eric M. Zolt. "Technology and Taxation in Developing Countries: From Hand to Mouse." *National Tax Journal* 61, no. 4 (2008): 791-821. Accessed November 27, 2020. <http://www.jstor.org/stable/41790481>.

Umoru D, Anyiwe MA. Tax structures and economic growth in Nigeria: Disaggregated empirical evidence. *Research Journal of Finance and Accounting*. 2013;4(2):65-79.

“‘WeBOC’ Clearance System Launched at PQ.” 2011. DAWN.COM. June 3, 2011. <https://www.dawn.com/2011/06/03/weboc-clearance-system-launched-at-pq/>.

Zafar T. Taxation System of Pakistan and its Impact on Economy. Punjab Economic Research Institute, Planning and Development Department. Lahore: Government of Punjab, Pakistan. 2017 Feb.